MEASUREMENT OF SPECTRA OF CHARGED HADRONS AND WEAKLY DECAYING STRANGE PARTICLES AT LHC ENERGIES WITH THE CMS DETECTOR

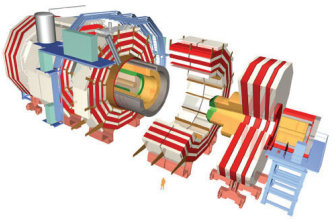
Ph.D. Thesis

Krisztián Krajczár

Doctoral School of Physics
Head: Prof. Ferenc Csikor

Particle Physics and Astronomy Programme
Head: Prof. Ferenc Csikor

Supervisor: Dr. Gábor Veres
ELTE TTK, Department of Atomic Physics



Eötvös Loránd University, Budapest
MTA KFKI RMKI, Budapest

July, 2011

Table of Contents

1. Introduction

Throughout the written history there always were humans interested in the structure and laws of the surrounding world, in particular in the study of the compositeness of matter. After the era of mainly philosophical arguments pro and contra of compositeness, in 1789 the french scientific researcher Antoine Lavoisier discovered the "law of conservation of mass" and defined an "element" as a basic substance that could not be further broken down (by the methods of chemistry at that time). Later, in 1803 John Dalton used the concept of elementary substances to explain why elements always react in ratios of small integer numbers, known as the "law of multiple proportions". The size of such indivisible substances were not known until 1865 when Johann Josef Loschmidt measured the size of the molecules that make up air. In 1869, Dmitri Mendeleev published the first functional periodic table where elements were arranged by atomic number.

In 1897, Joseph John Thompson discovered the electron and concluded that they were components of every atom. Thus the belief in the indivisible atoms was demolished. Ernest Rutherford's 1911 analysis of the famous Geiger-Marsden experiment suggested that the atom has a positive core, nucleus, while the negative electrons circle this core. In 1919, Rutherford discovered that the hydrogen nucleus is present in other nuclei, which he named proton. In 1931, Walther Bothe and Herbert Becker found that energetic *alpha particles* impigning on certain light elements produce radiation. This radiation was explained by James Chadwick in 1932 with a proton-like but electrically neutral particle, the neutron.

The structure (if any) of protons were probed with high energy particle collisions (due to the de Broglie wavelength, high energy is necessary to obtain the required resolution to be able to "see" inside the proton). The experimental results implied that nucleons are also composite, they are built up by partons (proposed by Richard Feynman in 1969), later known as quarks and gluons. The experimental study of the properties of quarks and gluons via their "freezed out" forms (hadrons, mesons) are the forefront of particle physics since then. My work described in this

thesis is an other small step in the gathering of experimental facts in the field of high energy particle collisions, and the dynamics of quarks and gluons building up most of the visible matter around us.

The thesis is organised as follows. In the first section, Section 1, I summarize the main theoretical aspects of my measurements presented in the thesis. This includes the discussion of the various types of high energy particle collisions, and the theories of these collisions. The first section ends with introducing the particle spectra and discussing the importance of their measurements. Section 2 briefly introduces the CERN research laboratory and its LHC particle accelerator. It is followed by the introduction of the experimental tool of my measurements, the CMS detector, in Section 3. The fourth section, Section 4, describes the measurement methods of the spectra of unidentified charged hadrons. I developed the so-called pixel counting analysis method and contributed to the development of the so-called minimum bias tracking, as it is explained in the preface of that section. Section 5 describes the obtained results on the spectra of unidentified charged hadrons. During the measurements I performed the pixel counting analysis. The section also contains a discussion on the results, and displays their afterlife and effects on other measurements. The next section, Section 6 introduces the measurement methods of the spectra of identified weakly decaying strange particles. I developed the so-called high efficiency method, as it is also mentioned in the preface of the section. Section 7 presents my results on strange particle spectra, including a discussion on the results.

I also included two appendices. The first one, Appendix A describes a data driven method to estimate the fraction of diffractive events in data, which was used for systematic uncertainty studies. Appendix B presents a collection of plots showing various comparisons between simulated and measured distributions. The aim of this appendix is to convince the Reader that the data and the response of the detector elements are well understood.

For the theoretical discussion in the first section I adopt the $c = \hbar = 1$ *natural units*, but for the latter sections on experimental methods this convention is not followed.

1.1 Various types of high energy particle collisions

Particle collisions at high energies can be divided into various types based on the particular physical interaction happening during the collision:

- **Elastic collisions.** These collisions are governed by the Coulomb scattering, and are described by quantum electrodynamics (QED). During the collision no new particles are produced (although due to the resulting acceleration of charges photons are emitted ("Bremsstrahlung")).

- **Inelastic collisions** (illustrated in Fig. 1.1).

  - *Diffractive dissociations.* Particle production is due to Pomeron[1] exchange. Further sub-classes are distinguished:

    * Single diffractive dissociation (SD). Single-Pomeron exchange happens, one of the protons is dissociated the other one travels onward with very little modification on trajectory. Single-sided particle production is observed in the forward detectors.

    * Double diffractive dissociation (DD). Two-Pomeron exchange happens, both of the protons are dissociated. Particles are detected in both forward regions of the detector.

    * Central diffractive dissociation (CD). The two protons are dissociated and further particles are produced due to the fusion of two Pomerons.

  - *Non-diffractive (ND) collisions.* Process described by quantum chromodynamics (QCD); "color charge" exchange happens.

  It is worth to define an other collision class based on the previous definitions:

  - *Non-single-diffractive (NSD) collisions.* Those inelastic collisions which are not single diffractive.

Since my measurements have a sensitivity to all these types of collisions except the elastic ones, and I will frequently refer to the various collision types later, I give a brief introduction to their theory in the next section.

[1] *Virtual particles* having all zero quantum numbers, just as the vacuum has. A brief introduction to Pomerons is given in Section 1.2.1.

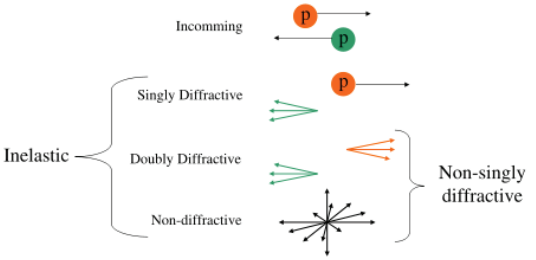


Figure 1.1: Sketch of the SD, DD and ND collision types. Details are given in the text.

1.2 The theory of the various collision types in a nutshell

This section is divided into two parts. In the first one the theory of diffractive collisions within the framework of the Regge theory is discussed. In the second part the theory of non-diffractive collisions, the QCD is introduced.

1.2.1 Regge theory and the Pomeron

The name Pomeron concerns the mechanism of diffractive processes at high energy. It originated from the Regge pole model developed in the sixties. Within this model one assumes that the dominant mechanism describing the high energy process is that of the Regge pole exchange.

Regge theory is the study of the *analytic properties* of scattering as a function of angular momentum, where the angular momentum is not restricted to be a real number ($\mathbb{R}$) but is allowed to take any complex value.

The general goal of the study of analytic properties of scatterings is to extract physics from the properties of the scattering matrix (S-matrix). The S-matrix is defined as the unitary matrix connecting asymptotic particle states in the Hilbert space of physical states: $S_{ab} = \langle b_{out}|a_{in}\rangle$. The elements in the S-matrix are known as scattering amplitudes. The S-matrix is Lorentz invariant and its poles in the complex-energy plane are identified with bound states, virtual states or resonances. Branch cuts of the S-matrix in the complex-energy plane are associated to the opening of a scattering channel.

The S-matrix can be written as

$$S_{ab} = \delta_{ab} + i(2\pi)^4 \delta^{(4)}(p_f - p_i) A_{ab}. \tag{1.1}$$

In case of the typical $2 \to 2$ scattering in particle accelerators, the A scattering amplitudes depend on the Lorentz invariant Mandelstam variables, s,t and u, which are defined as

$$\begin{aligned} s &= (p_1 + p_2)^2 = (p_3 + p_4)^2, \\ t &= (p_1 - p_3)^2 = (p_2 - p_4)^2, \\ u &= (p_1 - p_4)^2 = (p_2 - p_3)^2, \end{aligned} \tag{1.2}$$

where p_1 and p_2 are the four-momenta of the incoming particles and p_3 and p_4 are the four-momenta of the outgoing particles. The variable s is also known as the square of the center-of-mass energy (invariant mass), and t as the square of the momentum transfer. It can be easily shown that $s+t+u = m_1^2+m_2^2+m_3^2+m_4^2$. The S-matrix is an analytic function of Lorentz invariants with only those singularities required by unitarity (arises from causality).

Taking into account the unitarity ($SS^+ = \mathbb{1}$) of the S-matrix and Eq. 1.1, the imaginary part of the scattering amplitude can be written as

$$2\Im m A_{ab} = (2\pi)^4 \delta^{(4)}(p_f - p_i) \sum_x A_{ax} A^+_{xb}, \tag{1.3}$$

where a is the initial, and b is the final asymptotic particle state. In the special case of $b = a$, Eq. 1.3 establishes a connection between the forward scattering amplitude and the total cross-section,

$$2\Im m A_{aa} = (2\pi)^4 \delta^{(4)}(p_f - p_i) \sum_x |A_{ax}|^2 \simeq 2s\sigma_{tot}(a \to x), \tag{1.4}$$

where s is one of the Mandelstam variables (the square of the centre-of-mass energy) and σ_{tot} includes both elastic and inelastic scattering. Eq. 1.4 is known as the *optical theorem*.

It can be shown that the so-called *crossing symmetry* holds for the scattering

amplitudes:

$$A_{ab\to cd}(s,t,u) = A_{a\bar{c}\to\bar{b}d}(t,s,u) \tag{1.5}$$

In the following I focus on the $s \to \infty$ limit of amplitudes (*Diffractive* or *Regge limit*). The amplitude in the $a\bar{c} \to \bar{b}d$ cross channel process may be expanded in Legendre polynomials (*partial-wave equation*):

$$A_{a\bar{c}\to\bar{b}d} = \sum_{l=0}^{\infty}(2l+1)a_l(s)P_l(\cos\theta), \tag{1.6}$$

where θ is the centre-of-mass scattering angle, $\cos\theta = 1 + 2t/s$. By crossing symmetry

$$\begin{aligned} A_{ab\to cd}(s,t) &= \sum_{l=0}^{\infty}(2l+1)a_l(t)P_l(1+2s/t) = \\ &= \frac{1}{2i}\int_c dl(2l+1)\frac{a(l,t)}{sin\pi l}P(l,1+2s/t) = \\ &= \frac{1}{2i}\int_c \sum_\eta dl(2l+1)a^\eta(l,t)\frac{\eta + e^{-i\pi l}}{2}P(l,1+2s/t), \end{aligned} \tag{1.7}$$

where $\eta = \pm 1$ and the integration is done over a contour c which contains the singularities of the l angular momentum space. Contributions are only picked up from the poles in l-plane: for large s, the integral along c is zero.

In the $s \to \infty$ *Regge limit*

$$A_{ab\to cd}(s,t) \xrightarrow[s\to\infty]{} \frac{\eta + e^{-i\pi\alpha(t)}}{2}\beta(t)s^{\alpha(t)}, \tag{1.8}$$

where the assumption that all singularities in $a^\eta(l,t)$ are single poles at $l = \alpha(t)$. In general, there can be many cuts and poles but the one with the largest real part will dominate as $s \to \infty$. The $\alpha(t)$ function is called *Regge trajectory*.

According to the previous limit and the optical theorem (Eq. 1.4)

$$A(s,0) \xrightarrow[s\to\infty]{} \sim s^{\alpha(0)} \implies \sigma_{tot} \sim s^{\alpha(0)-1}, \tag{1.9}$$

where $s >> t$ (or $t \to 0$) form of the diffractive limit is taken. Note that if

the four-momenta of the incoming particles only suffer a small change in the interaction, it implies that $t \approx 0$. From the experimental point of view this limit means forward particle production, which is in good accordance with the experimental signals associated with diffractive collisions (Section 1.1). In the following we focus our attention on $\alpha(0)$.

The *Okun-Pomeranchuk Theorem* states that if the exchanged Regge trajectory $\alpha(t)$ does not have the quantum numbers of the vacuum, then $\alpha(0) < 1$. In addition the *Foldy-Peierls Theorem* implies that if $\alpha(0) \geq 1$, the Regge trajectory exchange must have the quantum numbers of the vacuum. Regge trajectories with $\alpha(0) \geq 1$ are called **Pomerons**. The Pomeron exchange is required by data since total cross-sections are observed to rise as $s \to \infty$ (cf. 1.9).

Diffractive collisions, described by Pomeron exchanges were observed by CMS [1]. The method of observation used in Ref. [1] is the same as the one developed for studying the systematic uncertainties due to diffractive event fraction of my measurements (see Appendix A for details about the diffractive event fraction in data).

1.2.2 Quantum chromodynamics

Quantum chromodynamics is the theory of strong interactions, interactions between quarks and gluons. It is a non-abelian quantum field theory with an SU(3) gauge symmetry group.

The Lagrangian of QCD is

$$\mathcal{L} = \bar{\Psi}_i(i\gamma^\mu\partial_\mu - m)\Psi_i - gG^a_\mu\bar{\Psi}_i\gamma^\mu T^a_{ij}\psi_j - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a, \tag{1.10}$$

where $\Psi_i(x)$ (i=1,...,6) are the *quark* fields in the fundamental representations of the gauge group, $G^a_\mu(x)$ (a=1,...,8) are the gauge fields (*gluons*) in the adjoint representation; they are functions of space-time. The γ^μ matrices are the Dirac matrices, T^a_{ij} are the generators of the SU(3) group. The Dirac matrices connect the spinor representation to the vector representation of the Lorentz group while the generators connect the fundamental, antifundamental and adjoint representation of

the SU(3) gauge group. $G^a_{\mu\nu}$ is the gauge field tensor which is given by

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f^{abc} G^b_\mu G^c_\nu, \tag{1.11}$$

where f^{abc} are the structure constants of SU(3). The constants m and g control the quark mass and coupling constant of the theory, subject to renormalisation in the full quantum theory. The QCD coupling strength is $\alpha_s \equiv g^2/4\pi$. The summation over index variables appearing twice in a single term is not written explicitly.

Now, consider a dimensionless physical observable, R, which depends on a single large energy scale, $Q >> m$ where m is any mass. Calculation of R as a perturbation series in the coupling $\alpha_s = g^2/4\pi$ requires renormalisation to remove ultraviolet divergences. This introduces a second mass scale μ – point at which subtractions that remove divergences are performed. Then R depends on the ratio Q/μ and is not constant. The renormalised coupling α_s also depends on μ.

But μ is arbitrary. Therefore, if we hold the bare coupling fixed, R cannot depend on μ. Since R is dimensionless, it can only depend on Q^2/μ^2 and the renormalised coupling α_s. Hence

$$\mu^2 \frac{d}{d\mu^2} R\left(\frac{Q^2}{\mu^2}, \alpha_s\right) \equiv \left[\mu^2 \frac{\partial}{\partial \mu^2} + \mu^2 \frac{\partial \alpha_s}{\partial \mu^2} \frac{\partial}{\partial \alpha_s}\right] R = 0. \tag{1.12}$$

Introducing

$$\tau = \ln\left(\frac{Q^2}{\mu^2}\right) \text{ and } \beta(\alpha_s) = \mu^2 \frac{\partial \alpha_s}{\partial \mu^2} \tag{1.13}$$

we have

$$\left[-\frac{\partial}{\partial \tau} + \beta(\alpha_s) \frac{\partial}{\partial \alpha_s}\right] R = 0. \tag{1.14}$$

This *renormalisation group equation* is solved by defining *running coupling* $\alpha_s(Q^2)$:

$$\tau = \int_{\alpha_s}^{\alpha_s(Q)} \frac{dx}{\beta(x)}, \; \alpha_s(\mu) \equiv \alpha_s. \tag{1.15}$$

Then

$$\frac{\partial \alpha_s(Q)}{\partial \tau} = \beta(\alpha_s(Q)) \text{ and } \frac{\partial \alpha_s(Q)}{\partial \alpha_s} = \frac{\beta(\alpha_s(Q))}{\beta(\alpha_s)} \tag{1.16}$$

and hence

$$R(Q^2/\mu^2, \alpha_s) = R(1, \alpha_s(Q)). \tag{1.17}$$

Thus all scale dependence in R comes from running of $\alpha_s(Q)$.

The running of the QCD coupling α_s is determined by the β function, which has the expansion

$$\beta(\alpha_s(Q)) = \frac{\partial \alpha_s(Q)}{\partial \tau} = -b\alpha_s^2(Q)[1 + b'\alpha_s(Q)] + \mathcal{O}(\alpha_s^4). \tag{1.18}$$

The b and b' parameters can be calculated and it turns out that the β function is negative overall[2] [3]. This means that as Q becomes large, $\alpha_s(Q)$ decreases to zero, QCD becomes a non-interacting theory. This feature of QCD is known as *asymptotic freedom*.

In the regime of small running coupling, perturbation computation techniques are usable. Corresponding to asymptotic freedom at high momentum scales, we have infrared slavery: $\alpha_s(Q)$ becomes large at low momenta (long distances). Important low momenta phenomenon is the *confinement*: partons (quarks and gluons) found only in colour singlet bound states, mesons and hadrons, size $\sim$ 1 fm. If we try to separate them it becomes energetically favourable to create extra partons from the vacuum (*fragmentation*). Partons produced in short distance interactions reorganise themselves to make the observed hadrons (*hadronisation*).

The experimental measurements at high momenta can be accurately described by perturbative QCD (see, for instance, Refs. [4, 5] for recent CMS results). However, at low momenta (due to large value of the running coupling) the equations of QCD cannot be solved by perturbative methods. There are various phenomenological models, which aspire to describe the low momenta measurements by applying

[2]David J. Gross, H. David Politzer and Frank Wilczek were awarded the Noble Prize in Physics in 2004 for their discovery of asymptotic freedom in the theory of the strong interaction [2].

certain assumptions on fragmentation and hadronisation. My measurements presented in this thesis probe this low momentum region, thus my results can only be compared to the predictions of such phenomenological models. After the next section, which introduces the particle spectra, I will return to the question of phenomenological models (Section 1.4).

1.3 Particle spectra

This subsection introduces the quantities which characterise the bulk properties of particles produced in collisions: the particle spectra. These quantities are the main subjects of my measurements presented in this thesis.

Defining the beam direction to be the z-axis, the energy and the momentum of a particle can be written as

$$E = m_T \cosh y,\ p_x,\ p_y,\ p_z = m_T \sinh y,\ \text{and}\ p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}, \tag{1.19}$$

where m_T, conventially called the *transverse mass*, is given by

$$m_T^2 = m^2 + p_x^2 + p_y^2, \tag{1.20}$$

p_{T} is the transverse momentum, and the rapidity y is defined[3] as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) = \ln \left(\frac{E + p_z}{m_T} \right) = \tanh^{-1} \left(\frac{p_z}{E} \right). \tag{1.21}$$

From Eq. 1.21 it is clear that y can only be computed for identified particles. However, the particle identification is a very challenging experimental task. A more convenient variable is the pseudorapidity, η. It is used to provide pure geometrical information on the produced particles without particle identification:

$$\eta = -\ln \tan \left(\frac{\theta}{2} \right) = \frac{1}{2} \ln \left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \right). \tag{1.22}$$

The pseudorapidity does not transform like a simple addition under longitu-

[3] In special relativity the definition of rapidity is a bit different; there $|\vec{p}|$ is used instead of p_z.

dinal boosts (as the rapidity does), but for relativistic particles ($|\vec{p}| >> m$) the pseudorapidity is approximately equal to the rapidity.

The differential cross section

$$E\frac{d\sigma}{d\vec{p}} = \frac{d^3\sigma}{d\phi dy p_{\mathrm{T}} dp_{\mathrm{T}}} \Rightarrow \frac{1}{2\pi p_{\mathrm{T}}}\frac{d^2\sigma}{dy dp_{\mathrm{T}}} = \frac{d^2\sigma}{\pi dy d(p_T^2)}, \tag{1.23}$$

where the arrow represents the average over ϕ, is called invariant because $d\vec{p}/E$ remains invariant under Lorentz transformation.

In QED, cross sections can be computed with high precision. However, in QCD this is not always the case, which makes the cross-section measurements harder to interpret (as already discussed in Section 1.2).

Experimentally, the differential cross-section can be measured in two ways: a) with 4π detector coverage or b) with the optical theorem (Eq. 1.4) based on forward scattering. Since the CMS detector does not provide a 4π coverage and is not equipped with the necessary forward detectors, thus measuring cross-sections is a hard and a not-so-well defined task. However, one can measure the yields of produced particles in a certain region of phase-space (given the limitations of the measurement apparatus). The quantity of particle yields is a well defined hadron level distribution.

The differential invariant yield can be expressed with the help of ϕ,y and p_{T} in the same way as we have already seen for the cross section in Eq. 1.23:

$$E\frac{dN}{d\vec{p}} \Rightarrow \frac{1}{2\pi p_{\mathrm{T}}}\frac{d^2N}{dy dp_{\mathrm{T}}}, \tag{1.24}$$

where N is the number of particles in the small $dy \times dp_{\mathrm{T}}$ phase-space box. Since the measurement of y requires particle identification (which is not always possible to perform), usually the measurement of pseudorapidity is performed instead (which is always possible). The invariant particle yield in this case can be calculated with the help of the Jacobian

$$\frac{d^2N}{dy dp_{\mathrm{T}}} = \frac{E}{|\vec{p}|}\frac{d^2N}{d\eta dp_{\mathrm{T}}}, \tag{1.25}$$

where

$$\eta(y) = \sinh^{-1}\left(\frac{\sqrt{m^2 + p_T^2}}{p_{\mathrm{T}}}\sinh y\right). \tag{1.26}$$

Note that this transformation is mass dependent. For the measurements of inclusive distributions the mass is taken to be that of the pion (since the majority of the particles produced in the collisions are pions). In order to reach a given y value, the measured quantities $d^2N/d\eta dp_{\mathrm{T}}$ have to be interpolated to $\eta(y)$ and transformed to $d^2N/dydp_{\mathrm{T}}$.

Having computed the $d^2N/dydp_{\mathrm{T}}$ ($d^2N/d\eta dp_{\mathrm{T}}$) distribution, the **particle spectra** dN/dy ($dN/d\eta$) and dN/dp_{T} can be obtained by integrating over p_{T} and y (η), respectively.

In general, the measurements of particle spectra are constrained both in y (η) and in p_{T} due to detector acceptance, reconstruction efficiency, and limitations in particle identification. Thus, for inclusive physics results extrapolation are necessary to the unmeasured kinematical region. The geometrical coverage of the experiments are usually limited to the central region corresponding to small $|\eta|$ (typically $|\eta| < 2.5$). In this limited region the (pseudo)rapidity distribution shows little (pseudo)rapidity dependence (as shown in Section 5 and Section 7). Thus, an extrapolation in (pseudo)rapidity would not be well constrained, and it would necessarily introduce large systematic uncertainties. However, depending on the experimental apparatus, the situation can be different for the p_{T} distributions. Several experiments (including CMS) is capable of measuring particles to very low-p_{T}, which gives a very powerful constraint for extrapolation.

Nevertheless, the extrapolation in p_{T} is not unique. There are several fit functions which can be used with differing applicability and theoretical support. The Boltzmann function is often used for "thermal" fits:

$$E\frac{d^3N}{dp^3}_B = A\exp\left(-\frac{E(p_{\mathrm{T}})}{T}\right), \tag{1.27}$$

where $E(y, p_{\mathrm{T}}) = \sqrt{m^2 + p_T^2}\cosh y$. Although it is widely used to describe the distributions of low-p_{T} particles produced in hadronic collisions [6, 7, 8, 9], it fails

to describe their high-p_{T} power-law tail. At high-p_{T} the empirical function

$$E\frac{d^3N}{dp^3}_{E} = \frac{A}{(1+p_{\mathrm{T}}/p_0)^n} \tag{1.28}$$

fits the data nicely [10, 11].

The Tsallis function [12] is the simplest generalization of the Boltzmann function by introducing the parameter q. (It is also called Levy function or q-exponential, with origin in non-extensive statistics.) The Tsallis function successfully combines and describes both the low-p_{T} exponential and the high-p_{T} power-law behaviours [13, 14, 15]:

$$E\frac{d^3N}{dp^3}_{T} = A\left(1+(q-1)\frac{E(p_{\mathrm{T}})}{p_0}\right)^{-q/(q-1)}. \tag{1.29}$$

In the low-p_{T} limit Eq. 1.29 is approximated by the Boltzmann exponential (Eq. 1.27) with $T = p_0/q$. At the high-p_{T} limit the previously mentioned empirical function (Eq. 1.28) is obtained with $n = q/(q-1)$. In practice, the following reparametrised form of the Tsallis distribution is particularly useful:

$$E\frac{d^3N}{dp^3} = \frac{dN}{dy}\frac{(n-1)(n-2)}{2\pi nT(nT+(n-2)m)}\left(1+\frac{E_T(p_{\mathrm{T}})}{nT}\right)^{-n}, \tag{1.30}$$

where the definition of the transverse energy is a bit different than previously: $E_T(p_{\mathrm{T}}) = \sqrt{m^2+p_T^2} - m$. This form provides both the inverse slope parameter T, characteristic for low-p_{T}, and the exponent n of the high-p_{T} power-law tail. The definition is such that the p_{T} integrated yield is simply dN/dy.

With the introduction of the $C(n,T,m)$ normalisation constant the Tsallis distribution can be written as

$$E\frac{d^3N}{dp^3} = \frac{1}{2\pi p_{\mathrm{T}}}\frac{E}{|\vec{p}|}\frac{d^2N}{dydp_{\mathrm{T}}} = C(n,T,m)\frac{dN}{dy}\left(1+\frac{E_T}{nT}\right)^{-n}. \tag{1.31}$$

As I will discuss it in Section 5 and Section 7, the Tsallis distribution is capable of describing my measured transverse momentum distributions accurately.

1.4 Importance of the measurement of particle spectra

High-p_{T} particle production can be calculated with perturbation theory of QCD due to the smaller values of the running coupling at that momentum region, as described in 1.2.2. The low-p_{T} end of the spectra can only be described by various phenomenological models which deal with fragmentation and hadronisation ("event generators"), or by calculable effective models (for instance, Refs. [16, 17]). These generators and models use experimental results to set up their parameters. As a consequence of the dominant production of low-p_{T} particles, I will compare my results to the predictions of various event generators (and in some cases to that of effective models) and not to the direct predictions of QCD. These event generators are the PYTHIA [18, 19] and the PHOJET [20, 21].

The event generators have various input parameters which affect their predictions. The procedure to find a set of parameters, with which certain physics observables can be described, are called "tuning". This tuning of the parameters happens using a selected number of observables at one or at a few energies [17, 22, 23, 24]. In general, it is true to say that tunes are made to describe a particular feature of the data and not all features at once. This results in multiple tunes even for a single event generator depending on the observable quantity a particular tune aspires to describe. The energy dependence of observables as predicted by the event generators is governed by other parameters which can also be re-tuned in the light of new measurements. Thus, experimental results in the field of theoretical description of particle production are of vital importance: a) to better understand the non-perturbative and b) to test QCD in the perturbative kinematical region.

The results also have experimental importance. The density of charged hadrons affects the expected lifetime of detector components exposed to radiation. It also affects the strategy of the recording of collision events (Section 4.1) and the preparations for collisions at higher energies. The *yield* of the charged hadron density measurement affects the amount of *background* of various other measurements, for instance, the measurement of spectra of strange particles presented in Section 6.

Hence, the measurement of particle spectra is truly the first step in discovering a new energy regime.

2. CERN and the Large Hadron Collider

2.1 CERN

The European Organisation for Nuclear Research (French: Organisation Européenne pour la Recherche Nucléaire), known as CERN, is the largest particle physics laboratory in the world, situated in the northwest suburbs of Geneva on the Franco-Swiss border. It was established in 1954. The organisation has twenty European member states, including Hungary (since 1992).

The main function of CERN is to provide the particle accelerators and other infrastructure needed for high-energy physics research. Numerous experiments have been constructed at CERN by international collaborations to make use of them.

Among the most successful experiments are UA1 and UA2. These experiments discovered the W and Z bosons, which together with a new technique for particle beam *cooling* lead to the Nobel Prize for physics being awarded to Carlo Rubbia and Simon van der Meer in 1984 [25]. These discoveries were achieved at the Super Proton Synchrotron (SPS) accelerator, which accelerated proton and antiprotons from 1981 to 1984. The physics programme of the next large accelerator, the Large Electron-Positron Collider (LEP), brought a large set of parameters measured very precisely by the ALEPH, DELPHI, OPAL and L3 experiments.

2.2 The Large Hadron Collider

The new accelerator of CERN, the Large Hadron Collider (LHC) is the world's largest and highest-energy particle accelerator. It is situated in a 3.8-metre wide concrete-lined tunnel, constructed between 1983 and 1988, which formerly housed the LEP. It crosses the border between Switzerland and France at four points, and most of its circumference lies in France.

The collider tunnel contains two adjacent parallel beam pipes that intersect at four points. The LHC accelerates protons (and also lead ions) in opposite directions around the ring. Beam-beam collisions happen at the four interaction points; the experiments are situated around these points. Some 1232 dipole magnets keep the

Figure 2.1: An aerial view of the vicinity of CERN. The large ring indicates the approximate position of the LHC, while the smaller one that of the SPS accelerator. The main CERN site is the triangular block just "below" the "intersections" of the rings. The Lake Geneva is at the top right corner of the picture, the International Airport Geneva is just "below" it.

beams on their circular path, while an additional 392 quadrupole magnets are used to keep the beams focused in order to maximize the chances of interaction between the particles in the intersection points. Since the magnets are superconducting magnets[4], which are needed to be kept at low temperatures (at 1.9 K, making them colder than the outer space), approximately 96 tonnes of liquid helium is used making the LHC the largest cryogenic facility in the world at liquid helium temperature. Given the large bending power of the magnets and the small curvature of the large accelerator ring, protons and lead ions can be accelerated to the highest energy collisions ever reached in controlled environment[5], up to a designed collision

[4]Electromagnets made from coils of superconducting wire. The magnetic flux can be much stronger than in ordinary iron-core electromagnets; no energy is lost to heat resistance.

[5]So-called *cosmic rays* might induce even higher energy collisions in the upper atmosphere or on the surface of the Moon.

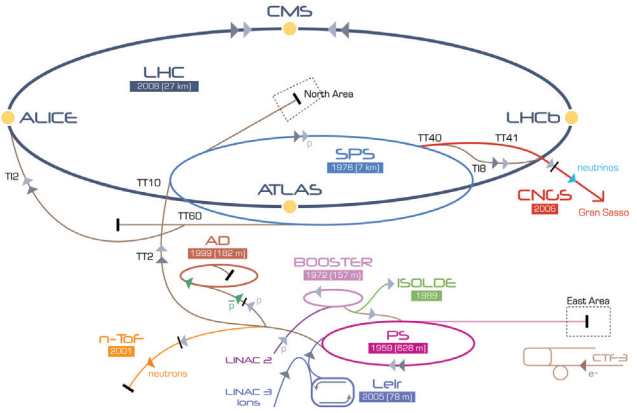


Figure 2.2: The accelerator (and decelerator) complex of CERN showing the LHC and its pre-accelerators.

energy per nucleon pair of 14 TeV.

The LHC is not a stand-alone accelerator, it is only the end-accelerator of the long accelerator-chain of CERN. Prior to the particles being injected into the LHC, they are prepared by a series of systems that successively increase their energy. The first system is the linear particle accelerator LINAC 2 generating 50-MeV protons, which feeds the Proton Synchrotron Booster (PSB). In the PSB the protons are accelerated to 1.4 GeV and injected into the Proton Synchrotron (PS), where they are further accelerated to 26 GeV. Finally, the SPS accelerator is used to increase their energy to 450 GeV before they are injected into the ring of LHC. In the LHC the proton bunches are accumulated and accelerated to their peak energy, and finally, collided.

Beams are not continual proton currents; the protons are organised in packages, called *bunches*. At designed operation conditions there will be 2808 bunches per beam, leading to collisions in every 25 ns. Each bunch will contain 1.15×10^{11} protons.

An important accelerator physics concept is that of the *luminosity*, $\mathcal{L}$. Luminosity is the factor of proportionality between the total cross section and the

collision rate: $dN/dt = \mathcal{L}\sigma$. It can be computed from basic accelerator physics quantities, $\mathcal{L} = fnN_1N_2/A$, where f is the revolution frequency, n is the number of bunches in one beam, N_1 and N_2 are the number of particles in each bunch and A is the cross section of the beam. The design luminosity in the LHC is $\mathcal{L}_D = 10^{34}$ $\mathrm{cm}^{-2}\mathrm{s}^{-2}$ [26].

At designed luminosity the total energy carried by the two beams in the LHC reaches 724 MJ (173 kilograms of TNT) and thus the beam dumps must absorb 362 MJ for each of the two beams. This is equivalent with the kinetic energy of a TGV train traveling with 150 km/h. The total energy stored in the magnets is 10 GJ (equivalent to 2.4 tons of TNT). These energies are even more impressive considering how little matter is carrying it: under nominal operating conditions (2808 bunches per beam, 1.15×10^{11} protons per bunch), the beam pipes contain 1.0×10^{-9} gram of hydrogen, which, in standard conditions for temperature and pressure, would fill the volume of one grain of fine sand.

Hadron colliders are well suited to the task of exploring new energy domains. The beam energy and the design luminosity of the LHC have been chosen in order to study physics at the TeV energy scale. A wide range of physics is potentially accessible with the seven-fold increase in energy and a hundred-fold increase in integrated luminosity over the previous hadron collider experiments.

The prime motivation of the LHC is to elucidate the nature of electroweak symmetry breaking for which the Higgs mechanism is presumed to be responsible. The experimental study of the Higgs mechanism can also shed light on the mathematical consistency of the Standard Model at energy scales above about 1 TeV. Various alternatives to the Standard Model invoke new symmetries, new forces or constituents. Furthermore, there are high hopes for discoveries that could pave the way toward a unified theory. These discoveries could take the form of supersymmetry or extra dimensions, the latter often requiring modification of gravity at the TeV scale. But the LHC also provides the possibility to study Standard Model physics at energies never measured before.

The LHC will also provide high-energy heavy-ion beams at energies over 30 times higher than at the previous accelerators, allowing to further extend the study

of QCD matter under extreme conditions of temperature, density, and parton momentum fraction.

The LHC started its physics programme in November, 2009, colliding two proton bunches at injection centre-of-mass energy, 900 GeV. In the first high statistics runs there were 4 bunches in each beam and the number of protons per bunches were 5×10^9 producing a luminosity of $\mathcal{L} \approx 2.21 \times 10^{26}$ cm^{-2} s^{-1}. The next milestone was the achievement of collisions at 2.36 TeV in December, 2009, with 2×2 bunches and a luminosity of $\mathcal{L} \approx 4.93 \times 10^{25}$ cm^{-2} s^{-1}. The collision rates were about 11 and 3 Hz, respectively. After the winter break, on 30 March 2010 the first proton-proton collisions happened at 7 TeV at a rate of about 50 Hz. Since then the LHC is kept colliding protons[6]. The luminosity is increasing, reaching 2×10^{32} with 1024 bunches per beam in June, 2011. The LHC already delivered an integrated luminosity of 1 fb^{-1} data by June, 2011, which was the plan for the full year of 2011.

There are seven experiments operating at LHC: a) four large: ALICE[7], ATLAS[8], CMS[9], LHCb[10], and b) three smaller: LHCf[11], MoEDAL[12], TOTEM[13]. Since there are only four collision points a few experiments are "watching" the same collision events (ATLAS-LHCf, CMS-TOTEM, LHCb-MoEDAL).

ATLAS and CMS are the two general purpose experiments; they are capable of measuring various types of particles (Standard Model and exotica) with high precision. The ALICE detector was primarily designed for lead-lead collisions, while the LHCb is specialized for b-physics[14]. Since my doctoral work is exclusively based on my work with the CMS detector, I denote a whole chapter for describing the most important design and detector elements of CMS (Chapter 3). It is necessary to better understand the measurements and the results discussed in later sections.

[6] With the exception of a short period of lead-lead collisions in Fall 2010.
[7] A Large Ion Collider Experiment
[8] A Toroidal LHC ApparatuS
[9] Compact Muon Solenoid
[10] The Large Hadron Collider beauty experiment
[11] Large Hadron Collider forward
[12] Monopole and Exotics Detector At the LHC
[13] TOTal Elastic and diffractive cross section Measurement
[14] Particularly aimed at measuring the parameters of CP violation – the violation of the simultaneous symmetry of charge and parity – in the interactions of b-hadrons (heavy particles containing a bottom quark)

3. The CMS Experiment

In this section I introduce the CMS detector with special emphasis on its components used for the measurements described in Section 4 and Section 6.

The Compact Muon Solenoid (CMS) detector [27] is a multi-purpose apparatus, which operates at the Large Hadron Collider (Fig. 3.1). CMS is installed about 100 metres underground close to the French village of Cessy, between Lake Geneva and the Jura mountains, around the so-called LHC Point 5 collision point.

The collision environment described in Section 2.2 requires a very careful design of the detectors. The total proton-proton cross-section at $\sqrt{s} = 14$ TeV is expected to be roughly 100 mb. For the currently achieved 7 TeV collision energy the inelastic cross section is measured to be 64 ± 2.4 (exp.) ± 6.9 (extr.) mb [28]. At design luminosity CMS (and ATLAS, too) will observe an event rate of approximately 10^9 inelastic events/s. This leads to a number of formidable experimental challenges. The online event selection process (online *trigger*) must reduce the huge rate to about 100 events/s for storage and subsequent analysis. The short time between bunch crossings, 25 ns (which corresponds to a distance of 7.5 metre), has major implications for the design of the readout and trigger systems.

At the design luminosity, a mean of about 20 inelastic collisions will be superimposed on each other (*pile-up*). This implies that $\mathcal{O}(1000)$ charged particles will emerge from the interaction region every 25 ns. The large frequency of collisions means that the products of an interaction may be confused with those from other interactions in the same bunch crossing. This problem clearly becomes more severe when the response time of a detector element and its electronic signal is longer than 25 ns. The effect of this pile-up can be reduced by using high-granularity detectors with good time resolution, resulting in low occupancy. This requires a large number of detector channels. The resulting millions of detector electronic channels require very good synchronization. The large flux of particles coming from the interaction region leads to high radiation levels, requiring radiation-hard detectors and front-end electronics. The detector requirements for CMS to meet the goals of the LHC

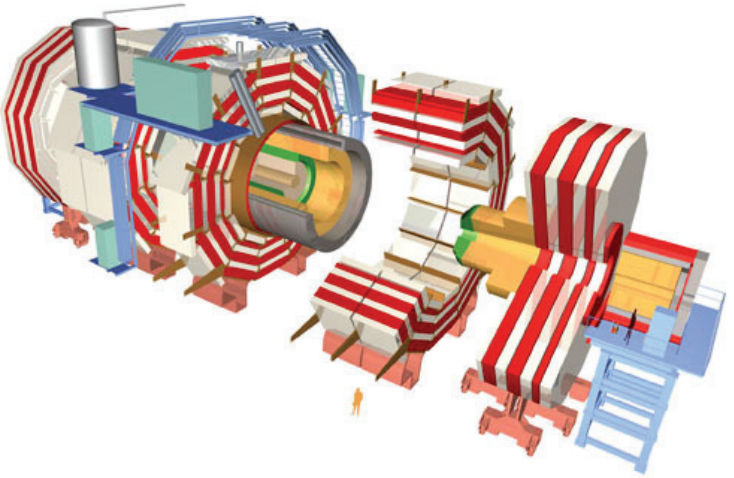


Figure 3.1: The CMS detector.

physics programme can be summarised as follows:

- **Tracking Detector.** Good charged-particle momentum resolution and reconstruction efficiency. Efficient triggering and offline *tagging* of τ's and b-jets, requiring high granularity *pixel detectors* close to the interaction region.

- **Electromagnetic Calorimeters.** Good electromagnetic energy resolution, good diphoton and dielectron mass resolution ($\approx 1\%$ at 100 GeV), wide geometric coverage, π^0 rejection, and efficient photon and lepton isolation at high luminosities.

- **Hadronic Calorimeters.** Good missing-transverse-energy and dijet-mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation.

- **Muon Chambers.** Good muon identification and momentum resolution over a wide range of momenta and angles, good dimuon mass resolution ($\approx 1\%$ at

100 GeV) and the ability to determine unambiguously the charge of muons with $p < 1$ TeV.

Thus, the main distinguishing features of CMS are a high-field solenoid, a full-silicon-based inner tracking system, and a homogeneous scintillating-crystals-based electromagnetic calorimeter. The coordinate system adopted by CMS has the origin centred at the nominal collision point inside the experiment, the y-axis pointing vertically upward, and the x-axis pointing radially inward toward the centre of the LHC. Thus, the z-axis points to the counter-clockwise beam direction. The azimuthal angle ϕ is measured from the x-axis in the $x - y$ plane, the radial coordinate in this plane is denoted by r. The polar angle θ is measured from the positive z-axis. Thus, the momentum transverse to the beam direction, denoted by p_T, is computed from the x and y components. The pseudorapidity η is defined as $\eta = -\ln\tan(\theta/2)$.

The various detectors of CMS and the *typical* "destiny" of various high-p_T particles are shown in Fig. 3.2. The journey of the particles are as follows.

- Charged hadrons cross the Tracker leaving signals in the detector and stop in the Hadronic Calorimeter.
- Neutral hadrons cross the Tracker unnoticed and stop in the Hadronic Calorimeter.
- Electrons and positrons leave hits in the Tracker and finally stop in the Electromagnetic Calorimeter.
- Photons traverse through the Tracker without leaving any trace and are detected by the Electromagnetic Calorimeter.
- Muons traverse through the whole detector on an S-shape trajectory. They are detected by the Tracker and the Muon Chambers.

In the following subsections I will introduce the subdetectors used for my measurements (Hadronic Calorimeter, the Beam Monitoring System, and the Tracker Detector).

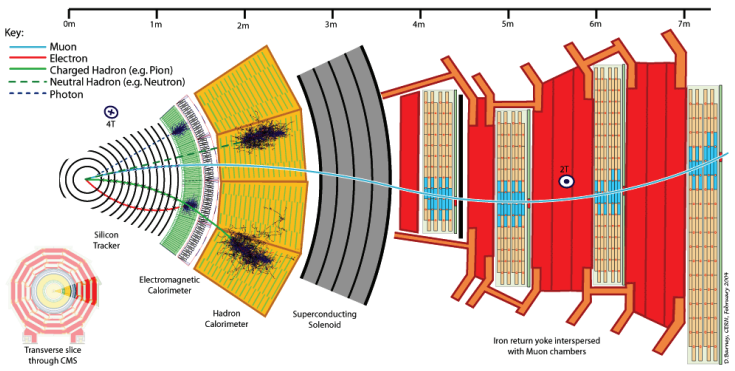


Figure 3.2: Schematic view of a section of the transverse segment of the CMS detector. The typical particle paths are also shown.

3.1 The Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) is a brass/scintillator sampling hadron calorimeter with coverage up to $|\eta| < 3.0$. The scintillation light is converted by wavelength-shifting fibres embedded in the scintillator tiles and channeled to photodetectors via clear fibres. This light is detected by photodetectors that can provide gain and operate in high axial magnetic fields. This central calorimetry is complemented by a *tail-catcher* in the barrel region ensuring that hadronic showers are sampled with nearly 11 hadronic interaction lengths. Coverage up to a pseudorapidity of 5.0 is provided by an iron/quartz-fibre calorimeter, the HF.

For the measurements presented in Section 4 and Section 6, I used the HF for triggering purposes (Section 4.1). Therefore, I describe it in more details.

3.1.1 Hadronic Forward Calorimeter (HF)

On average, 760 GeV per proton-proton interaction is deposited into the two forward calorimeters, compared to only 100 GeV for the rest of the detector. Moreover, this energy is not uniformly distributed but has a pronounced maximum at the highest rapidities. At $|\eta| = 5$ after an integrated luminosity of 5×10^5 pb^{-1} ($\approx$ 10 years of LHC operation), the HF will experience $\approx$ 10 MGy radiation dose

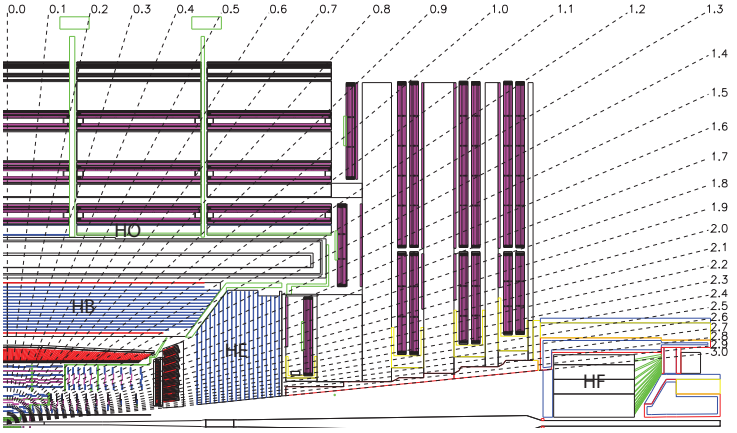


Figure 3.3: Longitudinal view of one quarter of the CMS detector showing the location of HF as well as the barrel (HB), endcap (HE) and outer (HO) hadronic calorimeters.

(for comparison, the average radiation dose affecting a human from an abdominal X-ray is 1.4 mGy). The charged hadron rates will also be extremely high. For the same integrated luminosity, inside the HF absorber at 125 cm from the beam-line, the rate will exceed 10^{11} per cm^2. In this hostile environment successful operation critically depends on the radiation hardness of the active material. This was the principal reason why quartz fibres were chosen as the active medium.

The forward calorimeter is essentially a cylindrical steel structure with an outer radius of 130.0 cm. The front face of the calorimeter is located at 11.2 m from the interaction point (the position of HF is indicated in Fig. 3.3). HF consists of a steel absorber structure that is composed of 5 mm thick grooved plates. Quartz-fibres are inserted in these grooves (the assembly of HF was partially done in MTA KFKI RMKI, Budapest). The detector is functionally subdivided into two longitudinal segments. Half of the fibres run over the full depth of the absorber (165 cm $\approx 10\lambda_I$) while the other half starts at a depth of 22 cm from the front of the detector. This arrangement makes it possible to distinguish showers generated by electrons and photons, which deposit a large fraction of their energy in the first 22

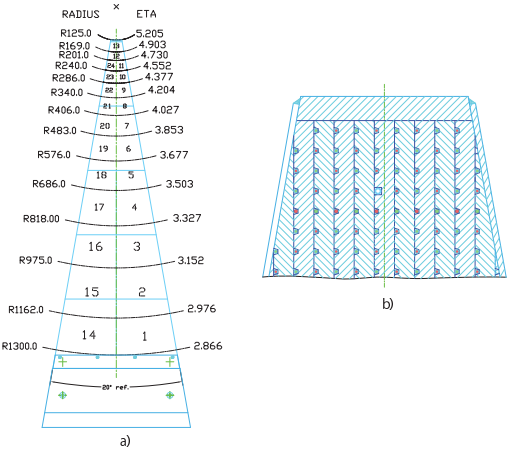


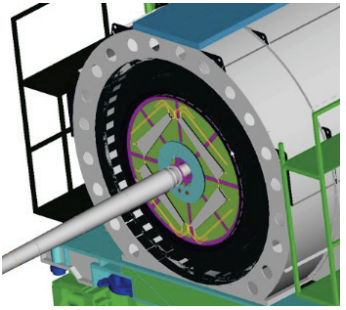
Figure 3.4: a) Transverse segmentation of a wedge showing the HF towers. b) An expanded view of the wedge shows the squared out groove holding the radioactive source tube.

cm, from those generated by hadrons, which produce nearly equal signals in both calorimeter segments on average. The calorimeter is azimuthally subdivided into 20° wide modular *wedges* (Fig. 3.4).

In the HF the signal is generated when charged shower particles above the Cherenkov threshold ($E \geqslant 190$ keV for electrons) generate Cherenkov light, thereby rendering the calorimeter mostly sensitive to the electromagnetic component of showers. Due to the use of Cherenkov light from quartz fibres, the HF is practically insensitive to neutrons and to low energy particles from the decay of activated radionucleids.

3.2 The Beam Monitoring System

There are a large number of sub-systems deployed with the primary purpose of monitoring radiation and passing beam bunches. It this section I restrict myself to describe two of them: the Beam Scintillator Counters (BCS) and the Button Beam Pickups (BPTX). Both of them played important roles in the recording of



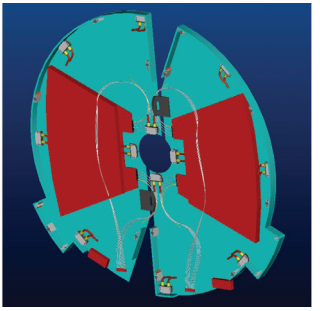


Figure 3.5: Layout of the Beam Scintillator Counters tiles. The *left*-hand panel shows the layout for BSC1 (mounted on HF), the *right*-hand panel for BSC2.

collision events (Section 4.1). Since both the BSC and the BPTX are sensitive to time structure below the 25 ns level; they can provide trigger inputs into the global CMS trigger.

3.2.1 The Beam Scintillator Counters

The Beam Scintillator Counters (BSC) are a series of scintillator tiles designed to provide hit and coincidence rates.

The detector consists of two parts. The BSC1 is located on the front of the HF, at ± 10.9 m from the interaction point (IP), and consists of two types of tiles. Next to the beampipe are the *disks*, segmented into 8 independent slices in ϕ, with an inner radius of 22 cm and an outer radius of 45 cm. The primary function of the disks is to provide the rate information corresponding to the beam conditions. In addition, there are four large area *paddles* further out, at a radial distance of between 55 cm and 80 cm. There is no functional difference between the disks and paddles. Both of them are capable of providing coincidence information which can be used to tag particles passing through the detector. The layout and the location of BSC1 can be seen in the left panel of Fig. 3.5.

The BSC2 is located at ± 14.4 m from the IP. It consists of two tiles on each side of the interaction point, with a minimum inner radius of 5 cm and a maximum outer radius of 29 cm (right panel of Fig. 3.5). The BSC2 can be used to tag bunches

with large out-of-beampipe activity (due to unintentional upstream collisions).

The area covered by the BSC is about 25% of the Tracker Detector; therefore these tiles are indicative of activity within bunch crossings.

3.2.2 The Button Beam Pickup

The BPTX is a beam pickup device specifically installed to provide the experiments with the timing structure of the LHC beam. This beam pickup is a standard button monitor used everywhere around the LHC ring for the beam position monitors. Two are installed for CMS: 175 m left and right upstream of the interaction point. At these locations there are two beampipes, the timing measurement is only done for the incoming beam.

The BPTX provides accurate information on the timing and phase of each bunch and its intensity. The phases of all the experimental clocks can be compared to the measured phase of each bunch with a precision better than 200 ps.

The signals from the BPTX are sent as inputs to the CMS global event recording logic. This provides flags on each bunch crossing as to whether: a) bunch in beam 1 is filled; b) bunch in beam 2 is filled; c) bunches in both beams are filled. The flag where both beams are filled indicates that collisions might occur in the bunch crossing.

3.3 The Tracker Detector

The tracking system of CMS is designed to provide a precise and efficient measurement of the trajectories of charged particles produced in the collisions, as well as a precise reconstruction of secondary vertices. The Tracker Detector surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m. The CMS solenoid provides a homogeneous magnetic field of 3.8 T over the full volume of the tracker. At the LHC design luminosity of 10^{34} cm^{-2} s^{-1} there will be on average about 1000 particles from more than 20 overlapping proton-proton interactions traversing through the tracker for each bunch crossing in every 25 ns. Therefore a detector technology featuring high granularity and fast response is required, such that the trajectories can be identified reliably and attributed to the correct bunch

crossing.

The intense particle flux will cause severe radiation damage to the tracking system. The main challenge in the design of the tracking system was to develop detector components able to operate in this harsh environment for an expected lifetime of 10 years. These requirements on granularity, speed and radiation hardness lead to a tracker design entirely based on silicon detector technology [29, 30].

The CMS tracker is composed of a pixel detector with three cylindrical (barrel) layers at radii between 4.4 cm and 10.2 cm and a silicon strip tracker with 10 barrel detection layers extending outwards to a radius of 1.1 m. Each system is completed by endcaps which consist of 2 disks in the *Pixel Detector* and 3 plus 9 disks in the *Strip Tracker* on each side of the barrel, extending the acceptance of the tracker up to a pseudorapidity of $|\eta| < 2.5$. With about 200 m^2 of active silicon area the CMS tracker is the largest silicon tracker ever built.

3.3.1 The Pixel Detector

The pixel system is the part of the tracking system that is closest to the interaction region. It contributes precise tracking points in $r-\phi$ and z and therefore is responsible for a small impact parameter resolution. With a pixel cell surface of 100×150 μm^2 emphasis has been put on achieving similar track resolution in both $r-\phi$ and z directions. As a result of this the precise reconstruction of the interaction point is possible.

The Pixel Detector covers a pseudorapidity range of $|\eta| < 2.5$. It consists of three barrel layers (BPix) with two endcap disks (FPix). The 53-cm-long BPix layers are located at mean radii of 4.4, 7.3 and 10.2 cm. The FPix disks extending from $\approx$ 6 to 15 cm in radius, are placed on each side at z=$\pm$34.5 and z=$\pm$46.5 cm (a schematic view of the Pixel Detector can be seen in Fig. 3.6). BPix (FPix) contain 48 million (18 million) pixels covering a total area of 0.78 (0.28) m^2. The arrangement of the 3 barrel layers and the forward pixel disks on each side gives 3 tracking points per charged particles over almost the full η-range.

For the barrel layers the drift of the electrons to the collecting pixel implant is perpendicular to the 3.8 T magnetic field of CMS. The resulting Lorentz drift leads

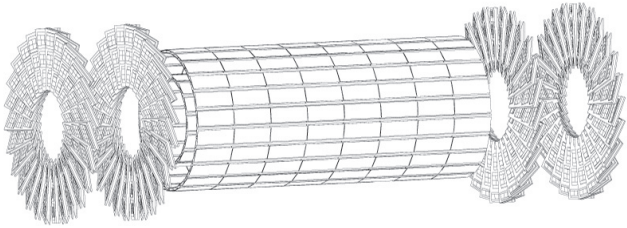


Figure 3.6: Schematic view of the Pixel Detector. It consists of three barrel and 2×2 endcap layers.

to charge spreading of the collected signal charge over more than one pixel. With the analog pulse height being read out a charge interpolation allows to achieve a spatial resolution in the range of $15-20$ μm. The forward detectors are tilted at 20° in a turbine-like geometry to induce charge-sharing. The charge-sharing is mainly due to the geometric effect of particles entering the detector at an average angle of 20° away from normal incidence; charge-sharing is also enhanced by the $E \times B$ drift. A position resolution of approximately 15 μm in both directions can be achieved with charge-sharing between neighbouring pixels.

Due to the small sizes of the pixel unit a particle traversing through the Pixel Detector deposits charges in multiple pixel cells. As a result such cells get above the readout threshold. Adjacent pixel cells above the readout threshold are grouped together into a higher level object, a *pixel cluster* (or *pixel hit*). The above-described scheme of pixel cluster creation is illustrated in the left panel of Fig. 3.7). Pixel clusters are two dimensional objects. Consequently, there is a connection between the size and shape of a cluster, and the local direction of the trajectory of the particle creating the cluster. The information contained in the shape of the clusters is exploited in my data analysis methods in various ways (details are given in Section 4).

3.3.2 The Strip Tracker

The silicon strip tracker is composed of 15 148 detector modules distributed among four different subsystems: Tracker Inner Barrel (TIB), Tracker Inner Disc

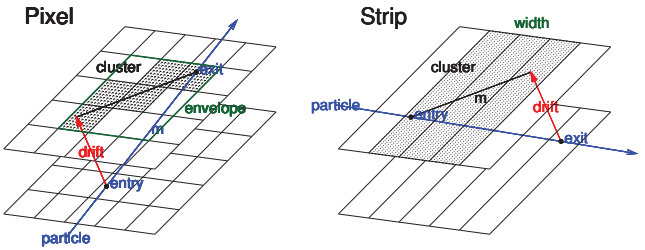


Figure 3.7: Schematic view of a pixel (*left*) and a strip (*right*) cluster. The blue arrow indicates the trajectory of a charged particle traversing a pixel layer (the thickness of the layer is indicated by two parallel surfaces). Due to the 3.8 T magnetic field the direction of the Lorentz drift plays a role in the development of the shape of the clusters.

(TID), Tracker Outer Barrel (TOB), and Tracker Endcap (TEC). The layout of the Tracker Detector (pixels and strips) is shown in Fig. 3.8. Each module carries either one thin (320 μm) or two thick (500 μm) silicon sensors from a total of 24 244 sensors. Depending on the geometry and number of sensors the active area of a module varies between 6 243.1 mm^2 (TEC, ring 1) and 17 202.4 mm^2 (TOB module). There are 512 or 768 silicon strip cells per modules.

The strip cells have elongated shapes compared to the pixels. The typical strip cell has a surface area of 100 μm $\times$ 10 cm. The *strip clusters* are the strip counterparts of the pixel clusters (right panel of Fig. 3.7). Due the elongated geometry of the strip cells the strip clusters are only one dimensional. This means that less information is contained in the shape about the local direction of the trajectory than in the case of the two dimensional pixel clusters. Still, even this smaller amount of information can be exploited as will be discussed in Section 4.

Tracker Inner Barrel (TIB) The TIB consists of four concentric cylinders placed at radii of 255.0 mm, 339.0 mm, 418.5 mm, and 498.0 mm respectively from the beam axis that extend from 700 mm to +700 mm along the z-axis. The two innermost layers host double sided modules with a strip pitch of 80 μm, while

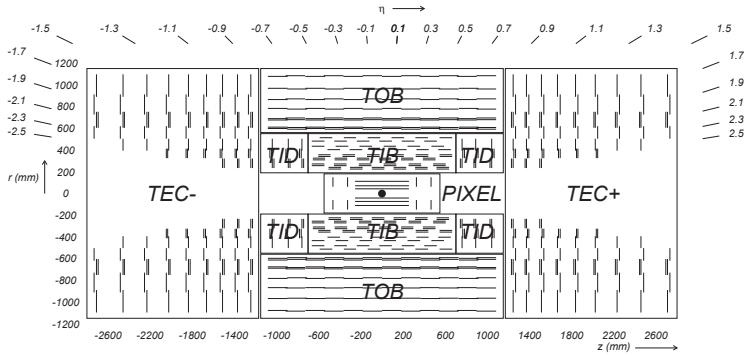


Figure 3.8: Schematic cross section of the CMS Tracker Detector. Both the pixel and strip layers are shown.

the outer two layers host single sided modules with a strip pitch of 120 μm. Fig. 3.9 shows the TIB layer ready for assembling in Pisa before transportation to CERN.

Tracker Inner Disc (TID) The TID± (at the ends of the TIB) are assemblies of three disks placed in z between ±800 mm and ±900 mm. The disks are identical and each one consists of three rings which span the radius from roughly 200 mm to 500 mm. The two innermost rings host back-to-back modules while the outer one hosts single sided ones. Together the full TIB/TID guarantee hermetical coverage up to pseudorapidity $|\eta| = 2.5$.

Tracker Outer Barrel (TOB) The TOB consists of a single mechanical structure (*wheel*) supporting 688 self-contained sub-assemblies, called *rods*. The wheel is composed by four identical disks joined by three outer and three inner cylinders. Each rod is supported by two disks, and two rods cover the whole length of the TOB along the z-axis. The wheel has a length of 2180 mm, and inner and outer radii of 555 mm and 1160 mm, respectively. The openings in the disks form six detection layers with average radii of 608, 692, 780, 868, 965, 1080 mm. Within each layer, the centres of gravity of the rods are displaced by ±16 mm with respect to the average

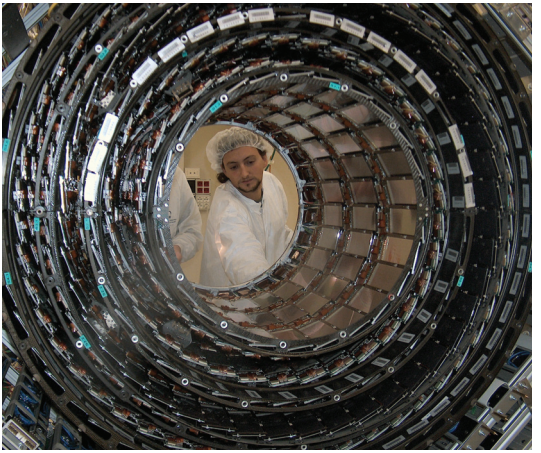


Figure 3.9: The TIB detectors are ready for assembling in Pisa before transportation to CERN.

radius of the layer, thus allowing for overlap in ϕ and therefore full coverage within each layer. The rod mechanics are designed in such a way to implement overlap of the silicon sensors at $z = 0$.

Tracker Endcap (TEC) The two TEC± endcaps extend radially from 220 mm to 1135 mm and from ±1240 mm to ±2800 mm along the z-direction. Each endcap consists of nine disks that carry substructures on which the individual detector modules are mounted plus an additional two disks serving as front/back termination.

3.3.3 Alignment of the Tracker Detector

The mounting precision of the elements of the Tracker Detector, $\mathcal{O}(100\mu m)$, was by far not sufficient for the goals of most physics analyses. This was one of my motivations to develop such an early measurement method, which can be applied to measure charged hadron densities without relying on the precise alignment of the Tracker Detector. This measurement approach has become the so-called pixel counting method (Section 4.3).

As the time of real collisions approached, the accuracy of the alignment was largely improved compared to $\mathcal{O}(100\mu m)$ using the reconstructed trajectories of cosmic ray muons. It was further improved using collision events, once they became available.

The alignment algorithm works as follows. It uses the hits produced by a traversing charged particle in the Tracker Detector. For every hit measurement i, position coordinates x_{hit} and corresponding errors are estimated within the local coordinate frames of the modules of the detector. Hit candidates are assembled into reconstructed particle trajectories, "track" candidates. Track parameters $\vec{q}$ for every track j are estimated by a track fit. This depends strongly on the alignment parameters $\vec{p}$. The alignment procedure uses the constraints implied by the track to estimate alignment corrections to the geometry. Deviations in geometry are reflected in the hit residual r, which is defined as the difference of the hit and the track prediction on the plane of the module, x_{track}, for each independent measurement coordinate,

$$r_{ji} = x_{ji,track}(\vec{p}, \vec{q}_j) - x_{ji,hit}. \tag{3.1}$$

The distribution of residuals normalized by their uncertainties is approximately Gaussian with a width of about 1, centred at 0, when there are no uncertainties in alignment parameters. Misalignment increases the spread of the residuals in general. This is reflected in an increase of the total χ^2-function, containing the goodness of all track fits,

$$\chi^2_{tot} = \sum_{j}^{tracks} \sum_{i}^{hits} \frac{r^2_{ji}}{\sigma^2_{ji}}. \tag{3.2}$$

The algorithms estimate alignment parameters by minimizing this function using millions of tracks. This needs sophisticated statistical approaches, since the track fits depend on the alignment. Two algorithms are applied. The local method named HIP (Hits and Impact Points) estimates the parameters for each module. Then iterations are needed to take the correlations with the track fit into account. It uses the same track model as the reconstruction. The global method (Millepede

II) fits all track and alignment parameters simultaneously. The advantage is that all correlations are considered, but its implementation in CMS is restricted to a helical track model.

The local and global methods deliver similar results and show remarkable improvement in the alignment quality. However, a combined approach gives the best results. The precision of the detector positions with respect to particle trajectories has been derived from the distribution of the median of the cosmic muon track residuals to be on average 3-4 μm RMS in the barrel and 3-14 μm RMS in the endcaps in the most sensitive coordinate [31]. This makes even the most sophisticated "tracking method" applicable on the first collisions (the tracking method is described in Section 4.5). However, this does not make the pixel counting method superfluous. First, in such of a such a complicated machine as CMS, which was built for 20 years, it is worth to be prepared for unexpected circumstances; and second, the pixel counting method still provides a largely independent physics result compared to that of the tracking method. The two methods have different sensitivity to the various sources of systematic uncertainties. Thus, they can cross-check each others results, and so they can highlight the problematic areas.

4. Measurement of spectra of unidentified charged hadrons

In this chapter I introduce my methods of measuring particle spectra in proton-proton collisions with CMS. First, I will start with the description of the event recording strategy. The final shape of the strategy did not emerge until the first real collisions. Of course, there were various studies based on simulated events, which laid the basis of the strategy, but it was partly modified in the light of real collisions. The strategy was developed by a small group of 7–8 people working on the analysis of the first collision data. I was one of the members of this group together with Ferenc Siklér (MTA KFKI RMKI), Gábor Veres (ELTE, Department of Atomic Physics), and colleagues from MIT. The discussion on event recording is followed by the introduction of the so-called pixel counting (or cluster counting) measurement method. This method was developed by me and Gábor Veres to measure the pseudorapidity distribution of charged hadrons [32, 33, 34, 35, 36, 37, 38]. I also performed this analysis on collision data; the results will be presented in the next section, Section 5. Then, the so-called tracklet method is introduced [35, 36, 37, 39]. This was developed entirely by my MIT colleagues, but was motivated by the pixel counting method. Indeed, it can be regarded as a coincidence version of pixel counting. The third method is the so-called tracking method [32, 33, 34, 35, 36, 37, 40, 41]. This was developed by me and Ferenc Siklér. It is capable of measuring both the pseudorapidity and the transverse momentum spectra. The section ends with a discussion on the systematic uncertainties affecting these analysis methods.

4.1 Triggering

My aim was to measure the inclusive distribution of pseudorapidity and transverse momentum of charged hadrons, which measurement (by definition) integrates the total cross-section of pp collisions. Thus the first task we face is what the optimal way of recognising pp collisions provided by the LHC is with *high efficiency* (that is to say with large acceptance probability) and *low fake rate* (only accepting real pp events). This mechanism of the recognition is called *triggering*.

In the CMS Experiment the triggering is done on two levels[15]. The first level is a selection entirely based on hardware signals and logical expressions between these signals; this is the *Level 1* trigger, or *L1*; such L1 triggers are among the so-called *online triggers*, which decide real time on the fate of a collision candidate. If the candidate is accepted by any of the L1 triggers, it has the potential to be *recorded*, otherwise it is discarded and lost forever. If the collision rate provided by the LHC is such that the triggering rate exceeds the *data to tape* bandwidth capabilities of the detector, triggers can be *prescaled*, that is to only accept every "n^{th}" collision, in order to reduce the load of the readout systems. Prescales can be applied for each trigger separately.

The second triggering step is the so-called *High Level Trigger* (or *HLT*). The HLT is a software which takes the L1 decisions and refines them. It is capable of running (even complex) reconstruction algorithms on data, thus imposing further selection criteria using higher level objects than raw hardware signal. The final decision of recording a collision candidate is taken by the HLT. If none of the HLT triggers accepts a candidate, it is lost and cannot be recovered. HLT triggers can also be prescaled independently of their L1 seeds.

After this triggering step further selection might still be need on the recorded data before a physical analysis. But since the data is already recorded, it can be done *offline*, at the time of data analysis. This further selection could have various reasons. One reason, for instance, is that of the limitations in the online computing resources, which limits the overall time we have at our disposal to decide on the of fate of a collision candidate. In the offline case such time constraints are much looser, thus we can impose more time-consuming selection requirements.

This makes the general triggering scheme look as follows. The first decisions about a collision candidate are made by the numerous independent triggers of L1, taking into account the possible L1 prescales. Then these decisions are fed to the HLT, which decides (taking into account the HLT prescales) whether the candidate

[15]This statement as well as my whole explanation of the triggering scheme is a simplification of the general triggering strategy. My aim here is to describe the importance and the main mechanism behind collision recording on the level which is necessary to understand my measurements. It is not intended to be a rigorous review of the triggering with CMS.

will be written to tape. If none of the HLT triggers accept it, the candidate is lost. Various physical analyses might impose additional requirements on the recorded data fitting the purpose of the analysis.

In the above paragraphs I consistently referred to the subject of trigger decisions as collision candidate. This is the usual case: usually we are interested in *pp* collisions, but this is not *always* true. We could be just as well interested in, say, cosmic muons, in which case the expression "collision candidate" is not correct. In the following I will more often use "event" to refer to the subject of the trigger decision.

Our aim was to record so-called minimum bias collision data. The term minimum bias refers to the fact that we try to minimize any kind of bias imposed on the recorded sample of collisions by the way they are selected or recorded. The ideal minimum bias data recording mode would collect all proton-proton collision events while it would be insensitive to any non-beam-beam collisions. However, in practice this is not attainable, but nevertheless, the goal is to maximize the efficiency and purity of event recording simultaneously.

CMS was not prepared for recording minimum bias data, its triggering strategy was geared towards sampling and recording rare events. The whole minimum bias triggering strategy had to be prepared from scratch.

4.1.1 Triggering the detector readout: online selection

For early collision data taking the CMS readout was triggered by a signal in any of the BSC segments, coincident with a signal from either BPTX indicating a beam or a bunch crossing the IP. Thus the detector collected data on any possible collision (both bunches of the two beams are filled: BPTX coincidence) and on beam background (only one bunch is filled: only one BPTX "fires") events coinciding with any BSC signal. These one-sided BPTX data are called "empty target" data, since no beam-beam collision is possible, and thus any recorded event must be classified as background of our intended measurement. The probability per bunch crossing during the early runs was well below one, so data only triggered by the BPTX would have contained a lot of empty events, events without collisions. Thus an

other detector must also be included in the trigger which was chosen to be the BSC. This was due to its high efficiency and low fake rate (see below).

It is essential to have precise information on the efficiency of BPTX and BSC as well as all the other triggers and event selection we apply since we have to correct for these effects in our measurements (our aim is to report results on inclusive hadron distributions).

The efficiency of the BPTX depends on the charge of the passing bunch. At the startup configuration the efficiency was measured to be higher than 99.99% above 10^9 protons/bunches. Since even during the first collisions there were 4×10^9 protons in each bunch, the BPTX was very much in the high efficiency plateau. As the number of protons increased in the bunches during the operation of LHC, the sensitivity threshold of BPTX was increased, but it was always kept at the efficiency plateau.

The efficiency of the BSC scintillator segments was measured based on the measurement of the MIP peak in the scintillators. Based on the MIP peak fit to the charge readout of the BSC, it was shown that the segments of BSC have on average a $> 95\%$ hit efficiency. This 95% efficiency was simulated in the reconstruction software by randomly erasing hits in the BSC segments.

4.1.2 Selecting events for analyses: offline event selection

The offline event selection is based directly on the information provided by the HF calorimeters. Events that contain a coincidence of at least one calorimeter tower with a total energy greater than 3 GeV in the positive and negative HF, i.e. on both sides of the IP, were selected. The efficiency of the event selection was checked as a function of the above-mentioned threshold energy in simulation and in data and the energy threshold was set to be above the noise level of HF. Fig. 4.1 shows the event selection efficiency as a function of the threshold energy for two event generators, PYTHIA [18] and PHOJET [20, 21], and for data. The efficiency in data was measured in so-called *zero bias* data requiring a reconstructed interaction *vertex*, that is the point of collision. Zero bias data are only triggered by the BPTX coincidence, it fires whenever two filled bunches cross each other in the detector. A vertex is required to

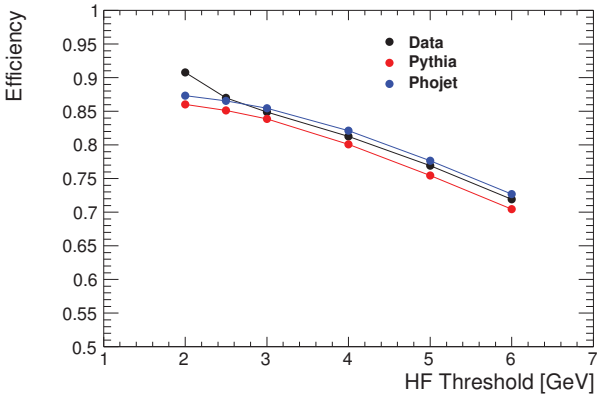


Figure 4.1: Event selection efficiency of the HF two sided coincidence trigger as a function of the hit energy threshold.

ensure that the event is not empty, it resembles a *pp* collision event. These events play the role of the denominator, while the fully selected events (described below) are the numerator. It can be seen that above 2.5 GeV a good agreement is achieved between the efficiency determined from data and from simulation.

In our offline event selection, the coincidence of BPTX signals were also required (in addition to this HF trigger) to largely suppress any random noise. The BPTX coincidence requirement does not reject any *pp* collision events, since the efficiency of BPTX is practically 100% (as already discussed).

The selected collision candidates still contain two kinds of background: i) beam halo[16] events and ii) so-called beam induced background events (see below). They are still among the selected events up to this point, since they can happen together with two filled bunches, or even with *pp* collisions. The beam halo events were rejected based on time difference of the hits in BSC1: by requiring two hits from the BSC1 stations on opposite sides of the interaction point to be within 20 ns of the mean flight time of 73 ns between them (the position of BSCs are described in Section 3.2.1).

[16]Secondary particles travelling with the bunch; they are usually muons from pion decays.

A further selection step was that of the requirement of a reconstructed interaction point or vertex, (the method of vertexing is described in Section 4.2).

Large-multiplicity beam background In some recorded events the tracker detectors were found to be filled with particle traces. The pattern of these hits were not compatible with a nominal collision in the centre of the detector and the number of the hits exceeded those observed in any beam-beam collision-like events. The origin of this background is still not accurately understood, but it seems so that these events originating from beam particles directly hitting some material upstream producing a spray of particles. Most of these background events are removed by the previous requirements on the HF energy deposits, but not all of them. Thus, a special filter was developed to remove this background completely.

As the primary signature of these high occupancy pixel events is the preponderance of extremely long clusters along the beam direction, this is the natural quantity for event-by-event discrimination. This is achieved by plotting all clusters according to their z-position and cluster size along the beam. Hits from primary tracks (tracks, produced in the collision) will leave a characteristic V-shaped pattern in this space. Non-primary hits (e.g. loopers[17], secondaries[18]) mostly fall outside this V-shape (left panel of Fig. 4.2).

To judge how compatible the primary vertex position is with the cluster-shape information, we define the *vertex uniqueness* as the ratio of clusters that fall within the V-shape to the average number of clusters inside the same V-shape when it is displaced by ± 10 cm along the beam. For events with only primary tracks and a properly reconstructed vertex, the vertex quality will be a large number. In the extremely high pixel multiplicity events, it will be close to unity. The two contributions corresponding to collisions and beam background are clearly visible in the correlation of vertex quality with pixel hit multiplicity in collision events (right panel of Fig. 4.2). This can be compared to Monte Carlo simulations where only collisions are present and empty-target bunch crossings where only background is

[17]Charged particles produced close to midrapidity with low-p_{T}; they are bent back by the magnetic field and cross the detector layers multiple times at a large angle leaving short clusters.

[18]Here it refers to any kind of particle not produced in the collision.

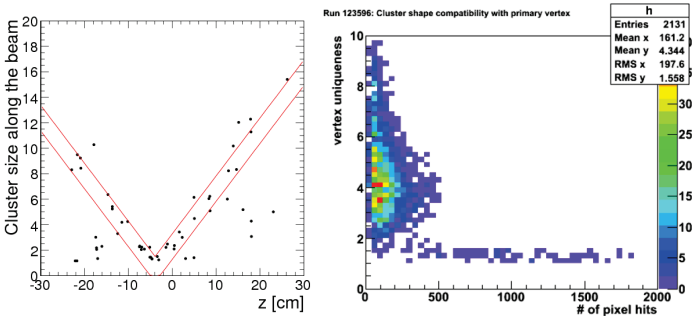


Figure 4.2: *Left*: The distribution of cluster size along the beamline as a function of the z positions, shown for a single event. The V-shaped correlation is highlighted with a band. *Right*: The vertex uniqueness versus the number of pixel hits in collision events. The BCS trigger, beam halo rejection and colliding bunches are already required.

present (Fig. 4.3). It is clear that a simple diagonal cut can be effective in cleaning up the sample in data without cutting much at all into the overall efficiency. These figures show the situation in the 0.9 TeV collisions. At 7 TeV the same could be said, although there, due to the increased average multiplicity, the precise position of the cut is different.

The diagonal was constructed such a way that less than 1-in-1000 event would be removed in simulation. Still it provides a powerful rejection of background in data (the fraction of surviving large-multiplicity beam background events is at the per mille level at 0.9 and 2.36 TeV, and it is less than 2×10^{-5} at 7 TeV).

Overall event selection efficiency The overall trigger and event selection efficiency for NSD collisions as a function of multiplicity as well as the multiplicity distributions in various event generators after event selection at 0.9 and 2.36 TeV are shown in Fig. 4.4. Again, very similar curves could be shown for the 7 TeV collisions. These efficiencies are obtained with the PYTHIA and PHOJET simulations in the following way. The denominator is the number of all generated events while the numerator is the number of selected events, for each multiplicity bin.

The fractions of various collision types in our event generators as well as our

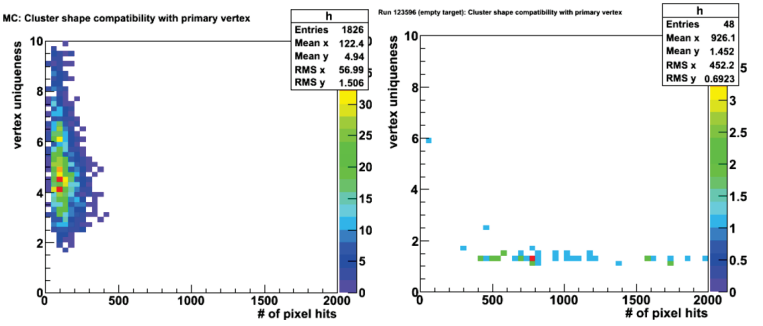


Figure 4.3: The vertex uniqueness versus the number of pixel hits in collision events in simulation (*left*) and in the empty target data (*right*).

	PYTHIA [%]						PHOJET [%]			
	0.9 TeV		2.36 TeV		7 TeV		0.9 TeV		2.36 TeV	
	Frac.	Eff.	Frac.	Eff.	Frac.	Eff.	Frac.	Eff.	Frac.	Eff.
SD	22.5	16.1	21.0	21.8	19.2	26.7	18.9	20.1	16.2	25.1
DD	12.3	35.0	12.8	33.8	12.9	33.6	8.4	53.8	7.3	50.0
ND	65.2	95.2	66.2	96.4	67.9	96.3	72.7	94.7	76.5	96.5
NSD	77.5	85.6	79.0	86.2	80.8	86.3	81.1	90.5	83.8	92.4

Table 4.1: Expected fractions ("Frac.") of SD, DD, ND and NSD processes obtained from the PYTHIA and PHOJET event generators before any selection, and the corresponding selection efficiencies ("Eff.") determined from the MC simulation, in percentage.

event selection efficiencies for these collisions are listed in Table 4.1. The table shows that the single diffractive events are largely suppressed by the event selection while the efficiency is very high for non-diffractive collisions. It can be computed from the table that (for instance) at 7 TeV the triggered 74.9% of inelastic collisions consist of 65.7% ND, 4.3% DD and 5.1% SD. Thus the correction of the measured distributions results in a correction of 25.1% (33.5% of 74.9) for inelastic, 5.1% (8.6% of 74.9) for NSD and 5.1%+4.3%=9.4% (12.6% of 74.9) for ND collisions. One can see that the correction is the smallest if the results are corrected for NSD collisions. This is the main reason behind our decision to correct the measurement for NSD collisions.

Although, I gave one reason of choosing to report NSD results, there were

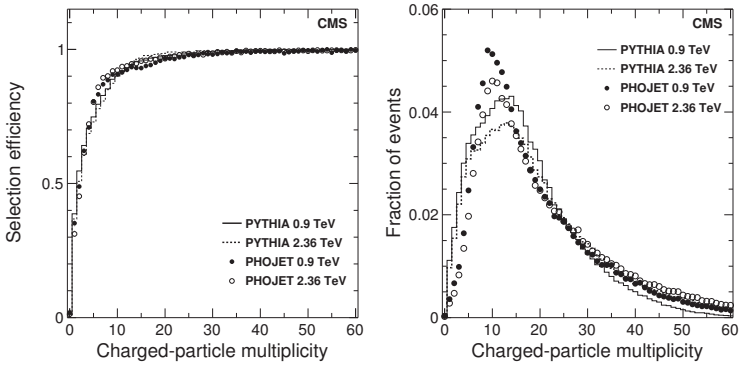


Figure 4.4: *Left:* The event selection efficiency expected for NSD events from the PYTHIA (histograms) and PHOJET (symbols) event generators as a function of generated charged hadron multiplicity in the region $|\eta| < 2.5$. *Right*: Generated multiplicity distributions of primary charged hadrons in the range $|\eta| < 2.5$ for $\sqrt{s} =$ 0.9 TeV (solid dots and histogram) and 2.36 TeV (open circles and dashed histogram) after the event selection is applied to the reconstructed events, using inelastic events from the PYTHIA (histograms) and PHOJET (symbols) event generators.

various other reasons in favour and against this. I would like to briefly sum up the arguments.

In favour:

- Our minimum-bias event selection criteria (described in details above) has a non-perfect selection efficiency for all collision types. The correction is the smallest in case of NSD collisions.

- Experimentally the SD content of the selected data can be largely reduced by applying certain event selection criteria (based on the topology of produced particles in these collisions: production in the "forward" region) which only slightly affects the ND content (such as the HF coincidence trigger). Furthermore, ND and DD are hardly distinguishable experimentally, thus including it in the measurement (ND+DD=NSD) reduces the total systematic uncertainties.

- Correcting the results for NSD collisions was chosen by a lot of experiments in the past [42, 43, 44, 45, 46, 47] and thus by reporting NSD results we gain the possibility of comparing them with those of the previous measurements.

- Although by correcting to true NSD events we directly make use of the description of SD event by the event generator used, the systematic uncertainties introduced here can be evaluated by using different models, different event generators. It can also be probed by data driven techniques (Appendix A).

- Theoretical comparisons from other models than Monte Carlo generators can only be made with results corrected to a "real" collision type, such as NSD or any other listed previously, and not to any *operational* event class definition, such as a certain number of particles in a certain phase-space, etc. Also see the first bullet of the against arguments.

  Against:

- A sample of pure NSD collision events cannot be separated from the rest of the collisions, since the particle production mechanism is not accessible directly. The directly accessible experimental information is on the level of hadrons and not on the level of partons. Thus instead of correcting to "true" NSD events, one should rather apply some hadron level cuts (thus defining a new experimentally derived collision category) and report results according to this definition, without correcting to any "real" collision type. This way the influence of the particular models of various processes within the event generators to the experimental results can be minimized (since the hadron level definition makes the collision category independent of the underlying collision model).

Based on the argument above the decision was made within the collaboration to correct the results for NSD collisions (the afterlife of this decision will be highlighted in Section 5.4).

4.2 Reconstruction of the interaction point

Information on the position of the interaction point (*primary vertex*, *vertex* or *PV*) is exploited by all analyses. In this section this method of vertex determination is described together with the determination of the so-called luminous region, or *beamspot*. This is the region where protons of both beams interact. The position and the shape of the luminous region is only used in the transverse momentum analysis. A few notations (such as triplets, etc.) are used here before they would be actually introduced in an ordered manner, which is done in the section on track reconstruction, Section 4.5. This "reversed" discussion has the benefit that once the vertex reconstruction is described, the pseudorapidity and transverse momentum measurements can be discussed in the order of increasing complexity.

The x and y positions of the luminous region is obtained from three-dimensional fits based on tracks reconstructed with $p_{\mathrm{T}} > 0.9$ GeV/c (track reconstruction is described in Section 4.5). The large p_{T} threshold does not introduce any bias here and has the advantage of suppressing the effect of multiple scattering on particle trajectories thus allowing a more precise determination of the luminous region. The RMS of the beamspot in both transverse directions was found to be less than 500 μm at 0.9 and 2.36 TeV, and 100 μm at 7 TeV. In my measurements no special vertex finding was done in the transverse direction, I only used the information provided by the above-mentioned constraints.

The situation, however, is different in the longitudinal direction. The RMS of the vertex distribution along the beam line is significantly larger (by three orders of magnitude) than in the transverse direction. The z position and the shape of the interaction region is obtained using the event-by-event primary vertices reconstructed as described below.

Due to the low collision rate during the recording of the analyzed data, the probability for more than one inelastic collision to occur in the same bunch crossing was less than 2×10^{-4} at 0.9 and 2.36 TeV, and 5×10^{-3} at 7 TeV. Thus, we can optimize the vertex reconstruction towards the reconstruction of a single interaction.

To reconstruct the z coordinate of the PV for each event, tracks consisting of triplets of pixel hits are formed. The minimal transverse momentum of these tracks

is $\approx$ 75 MeV/c. The tracks are required to originate from the transverse vicinity of the beamspot with a transverse impact parameter (d_T) smaller than 2 mm. Of these, only tracks with $d_T < 4\sigma_T$, where σ_T is the quadratic sum of the uncertainty of d_T (determined from the position measurement errors and from uncertainties due to multiple scattering) and the RMS of the beamspot in the transverse direction, are used in the vertex reconstruction. This is to reject tracks, which are likely not produced in the primary interaction, from the determination of the primary interaction point.

The vertex-reconstruction algorithm uses the z coordinate of the tracks at the point of closest approach to the beam axis and the corresponding estimated measurement uncertainty, σ_z. It performs an agglomerative clustering by adding tracks to form groups. These groups (denoted the i^{th} and j^{th} group) are then merged based on their normalized distance, $d_{ij}^2 = (z_i - z_j)^2/(\sigma_i^2 + \sigma_j^2)$ where σ_i and σ_j are the uncertainties of the z_i and z_j positions, with a fast nearest-neighbor search algorithm [48]. The z position and its uncertainty σ_z for the newly joined group are calculated using a weighted average. The clustering process stops when the smallest normalized distance, d_{min}, between the remaining groups gets larger than 8 at 0.9 and 2.36 TeV, and 12 at 7 TeV. These threshold values of d_{ij} were optimized using simulated events. The objects of optimization were the reconstruction efficiency[19] (aimed to maximize), the fake vertex fraction[20] (aimed to minimize), and the split vertex fraction[21] (aimed to minimize). Only vertices formed from at least two tracks are considered further except when there is only one track reconstructed in the event. In case of one-track-vertices the PV position is given by the point of closest approach of the track to the beam axis.

The fraction of single-track vertices in the data sample is 1.7% at 0.9 TeV, 1.3% at 2.36 TeV, and 0.9% at 7 TeV. The overall PV reconstruction efficiency, evaluated after all other event selection cuts are applied can be evaluated in data. This efficiency is just the fraction of selected events (without requiring the vertex,

[19]The fraction of found vertices.

[20]A vertex is fake if it has no associated simulated partner.

[21]The fraction of vertices which are reconstructed multiple times. It usually manifests itself as two vertices close to each other, which are associated to the same simulated vertex.

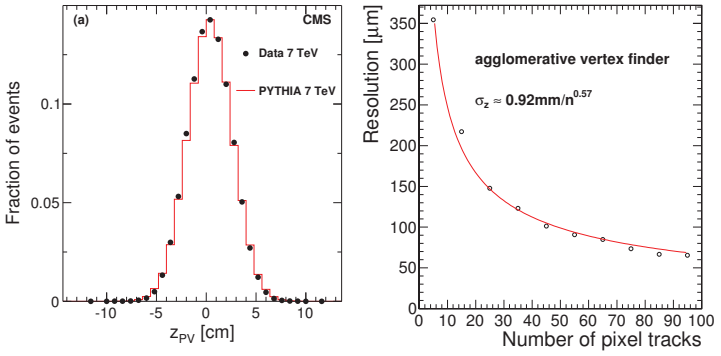


Figure 4.5: *Left*: The distribution of the reconstructed z position of the primary vertex in the 7 TeV data (symbols), compared to the same distribution from Monte Carlo simulation (lines). *Right*: Resolution of the vertex finder in 7 TeV PYTHIA simulation using the ATLAS tune [49].

but expecting a reconstructed track in the event) which have a vertex; the efficiency was found to be 99.2%. The probability of reconstructing more than one primary vertex candidate is 5.0% at 0.9 TeV, 7.4% at 2.36 TeV, and is below 6.0% at 7 TeV due to the increased d_{min}. When more than one PV candidate is reconstructed, the vertex composed of the largest set of tracks is chosen. The resolution of the primary vertex reconstruction was obtained from simulation. It is a function of the associated track multiplicity (m_{trk}) and can be parameterized as 0.87 mm $/{m_{\text{trk}}}^{0.6}$ at 900 GeV, and as 0.92 mm $/{m_{\text{trk}}}^{0.57}$ at 7 TeV.

The distribution of the reconstructed PV positions along the beam axis is shown in the left panel of Fig. 4.5. It is compared with that obtained from the simulation, which was adjusted to match the measured beamspot in three dimensions. All measurements presented in this thesis rely on the vertex being reconstructed as described above.

4.3 The pixel counting method

This method is based on the correlation between the number of hits in the barrel layers of the Pixel Detector and the number of produced particles in the collision. Although, the correlation is not precise, various cuts and corrections can be performed to suppress and subtract the hits due to background processes.

The measurement is performed on each pixel barrel layer separately, without using any information from other layers. Thus, the method provides three largely independent results, which can be combined. Using a simple hit counting technique, instead of a more complex procedure (for instance, the tracking method as described in Section 4.5), leads to a much better access to low-p_{T} particles (smaller minimal p_{T}) and is less sensitive to detector misalignment (about the alignment of the tracker see Section 3.3.3). However, the particles to be detected are still required to be energetic enough (more than 30 MeV/c transverse momentum) to reach the first pixel layer. This threshold is not only set by the magnetic field (which would make the particle trajectories to curve back before reaching the sensible detector layer), but also by the 0.8 mm thick Berilium beam pipe which is to be traversed.

Since this analysis deals with single layers, we have limited room to directly measure corrections from data. In some cases (loopers), we can do this (Section 4.3.2). In case of secondaries, we have to rely upon simulation. Because of this, the sensitivity for the properties of simulation was checked carefully. I prepared various distorted simulations to be used as "data" and performed the analysis using the original simulated sample as "simulation". Such distortions included changing the multiplicity density by a factor of two, and rescaling the transverse momentum of simulated particles by a factor of two. Even with these extreme modifications the method was able to reconstruct the true multiplicity distributions within $\sim 6\%$. The actual simulation I used for the analysis of real data was much closer to the data than the above-mentioned factors of two. The systematic uncertainties affecting this analysis method are discussed in details in Section 4.6.

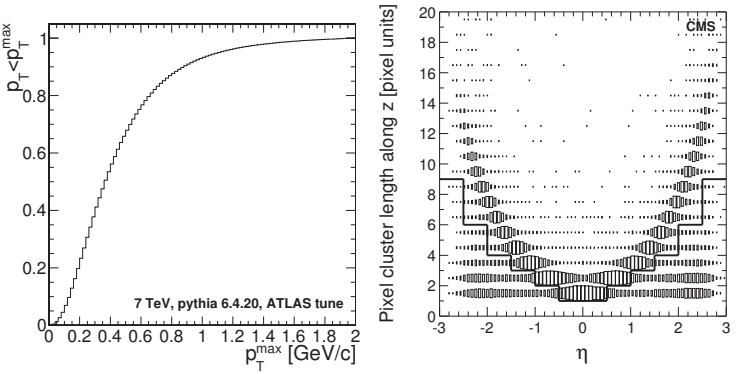


Figure 4.6: *Left*: Fraction of particles below p_T^{max} in the PYTHIA event generator as a function of p_T^{max}. *Right*: Pixel cluster size along the beam line as a function of η in data after all event selection is performed. The solid line shows the cluster shape cut applied in the analysis.

4.3.1 Cluster selection

A large fraction of the background hits originating from non-primary tracks can be rejected by cutting on the size of their clusters, or on the energy deposited in the clusters as a function of their pseudorapidity. The reason of this, and the precise way to apply these features in the analysis is explained in this subsection.

The value of η is calculated from the position of the hit relative to the primary vertex (determined as described in Section 4.2), $\eta = -\ln(\tan(\theta/2))$, where θ is the polar angle, with respect to the beam, of the straight line that connects the primary vertex and the cluster.

In case of particles with large p, or of zero CMS magnetic field, the η of a primary cluster can be easily computed from the z position of the cluster and the radius r of the layer it sits in: $z = r\sinh(\eta)$. However, in the non-zero homogeneous magnetic field of CMS, the particles move on a helical trajectory, if multiple scattering and energy loss are neglected. Then the relation between the above quantities are modified. The $\Delta\eta$ difference between the correct η', calculated with the helical trajectory, and the apparent η, calculated with a straight line connecting the

primary vertex and the hit can be estimated:

$$\ln\left[\frac{p_{\mathrm{T}}}{0.6r}\arcsin\left(\frac{0.6r}{p_{\mathrm{T}}}\right)\right] = \ln\left(\frac{\sinh(\eta')}{\sinh(\eta)}\right) > \eta' - \eta = \Delta\eta, \tag{4.1}$$

where p_{T} is in GeV/c and r is in meters. The right hand of the equation becomes approximately equal (instead of being larger) to the left hand for large η. The resulting correction is rather small: for the first pixel layer, it is 0.2 units of η for $p_{\mathrm{T}} = 30$ MeV/c, 0.1 for 38 MeV/c, 0.05 for 51 MeV/c and 0.01 for 110 MeV/c. Due to the small magnitude of this correction, the fact that the $dN/d\eta$ distribution is rather flat in the detector acceptance, that large majority of the final state particles have larger transverse momentum than 100 MeV/c, and that the multiple scattering further complicates the relation, this effect was neglected.

Particles traveling at a small θ polar angle leave larger clusters in the silicon barrel layers due to their small crossing angle. It can be easily shown that the *cluster size* along the beamline $\propto \sinh(\eta)$ while *cluster charge* $\propto \cosh(\eta)$. The cluster size as a function of pseudorapidity is shown in right panel of Fig. 4.6). Particles from background processes often have smaller clusters and smaller charge, since their crossing angle is not correlated to the η of the hit. Thus, we can reject clusters from background processes by cutting on the cluster size or cluster charge.

Both the size and the charge information can be used to perform the analysis, it is a matter of choice. Since the precise pixel charge response of the detector had non-negligible uncertainties during early collisions, the cluster size based analysis was performed on the data. The cluster size is only a *binary* information regarding the charge: the pixel charge is above or below the detector readout threshold; the precise value of the charge plays no role, hence the precise charge calibration is less important.

The position of the cluster selection cut is defined in the η bins of the measurement by minimizing the sum of the number of primary clusters below and background clusters above the cut. On Fig. 4.6 the cut is denoted by the horizontal black lines. The effectiveness of the cut in reducing the background is the largest at high-η, and it becomes 0 at $\eta \approx 0$.

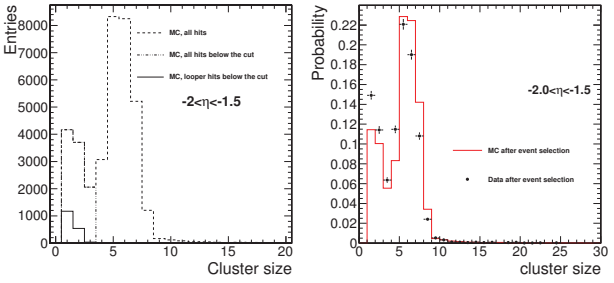


Figure 4.7: *Left*: Cluster size distribution in the $-2 < \eta < -1.5$ region from Monte Carlo simulations. *Right*: Cluster size distribution in the same η bin in data compared to simulations.

4.3.2 Correction for loopers

Charged particles with low transverse momenta ($p_{\mathrm{T}} < 800$ MeV/c) do not reach the calorimeters, and may move on a helical trajectory for more than half a turn. One turn on the helix takes $17.4 \times E$ nanoseconds, where the total energy of the charged particle, E, is measured in GeV. This formula gives 2.4 ns for pions and 16.4 ns for protons at low momentum. In reality the helix is not perfect, it is distorted due to the energy loss in the material traversed by the particle. Particles which cross a detector layer multiple times due to their bent trajectories are called loopers. Loopers can either be primary or secondary particles.

Ideally, only the first hit (that is produced during the first half "turn" of the looper) should be *counted* in the analysis, the other hits classify as background. A large fraction of the second, third, etc. hits produced by looper particles are eliminated by the cluster selection.

The fraction of late looper hits, among the clusters passing the cluster selection, can be estimated in a data driven way as follows. The idea is that the clusters rejected by the cluster selection contain information about the abundance of loopers. The left panel of Figure 4.7 shows the distribution of cluster size (along the beamline) in the $-2 < \eta < 1.5$ bin. This is a vertical slice of the cluster size as a function of η distribution, as it is shown in the right panel of Fig. 4.6, but per-

formed in simulated events. The figure shows that most of the looper contribution is in the rejected region. Still, the looper contamination of the selected clusters can be estimated taken the shape of the distribution of looper hits from simulation and scaling it with the relative amount of rejected clusters in data with respect to that in simulation. The right panel of Figure 4.6 compares the cluster size distributions in data and in simulation. Data shows a larger background peak than simulation; thus, a larger amount of loopers are accounted for than what is present in the simulation.

At small $|\eta|$ values the loopers (and any other kind of background) are not effectively discriminated from primary clusters by the cluster selection. In that region the looper correction is scaled up according to the observed excess of loopers at larger η. In the following (whenever it is not stated otherwise) I will use "selected clusters" as a shorthand for "clusters, which pass the cluster selection and have the looper hits removed".

4.3.3 Non-looper background

Clusters due to secondary particles cannot be removed completely by the cluster size cut. Thus it is necessary to correct for them. From simulation we can calculate the ratio of the η distributions of the reconstructed hits, H^{MC}, which pass our hit selection cuts and are still present after looper subtraction, and of the primary charged hadrons, P^{MC}, for a given M:

$$\chi(\eta, M) = \frac{H^{MC}(\eta, M)}{P^{MC}(\eta, M)} \tag{4.2}$$

The $\chi(\eta, M)$ function has to be evaluated by first collecting the numerator and the denominator separately from a large number of simulated events before dividing the two. This function is used to convert the $H(\eta, M)$ hit distributions into the corresponding primary hadron distributions.

$\chi(\eta, M)$ is not strongly physics model-dependent (as long as the secondary particles do not dominate), since it mainly contains information on the detector geometry (including the small rapidity correction between helical and straight trajectories). In a perfectly hermetic and 100% efficient detector, $\chi(\eta, M)$ will be slightly above unity, since not only the primary, but also the secondary particles can gener-

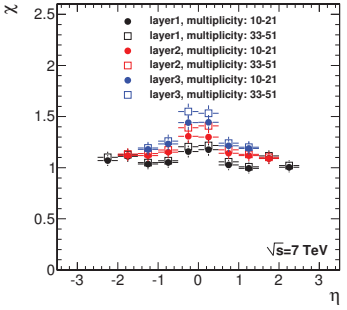


Figure 4.8: The χ correction for two multiplicity bins in the cluster size based analysis using 7 TeV PYTHIA ATLAS tune events.

ate a hit. For detectors covering a small solid angle, its values will be between 0 and 1. For very large multiplicities (typically in PbPb collisions), $\chi(\eta, M)$ may decrease with increasing M because of the more and more significant occupancy (provided the primary/secondary ratio stays roughly constant). However, the occupancy of the silicon pixel layers in p-p collisions is negligibly small; it remains around 1% even for the head-on PbPb collisions.

$\chi(\eta, M)$ is calculated for each barrel layer separately. For the outer layers $\chi(\eta, M)$ is larger than for the innermost ones due to the increased probability of a nuclear interaction with the material of the detector, and the larger amount of weak decays occurring in a larger volume. Figure 4.8 shows $\chi(\eta, M)$ evaluated in simulated PYTHIA events, for two multiplicity bins in the first, second, and third layer. The conversion factor in the first layer ranges from 1.0–1.2, and it becomes larger for layers further from the collision. The average value of the correction is 1.10, 1.23, and 1.41 for the first, second, and third layer, respectively. This average 1.10 correction has the following constituents: I subtract 6% for decay products, 8% for particles produced in interactions of a primary particle with the material of the detector, 3% for the overlapping detector layers; I add 7% for primaries which were rejected by the cluster size cut.

The pseudorapidity distribution of charged particles for a fixed multiplicity bin

can now be calculated from the measured $H(\eta, M)$ hit distributions, correcting with the hit/hadron ratio, and normalizing to the number of real data events passing the event selection, $E_{sel}(M)$:

$$\frac{dN_{sel}}{d\eta}(\eta, M) = \frac{1}{\chi(\eta, M)} \frac{H(\eta, M)}{E_{sel}(M)}. \tag{4.3}$$

The "sel" subscript in $dN_{sel}/d\eta$ refers to the fact these multiplicity distributions are not corrected for the event selection efficiency. They are only valid for the set of events actually selected. The selection-efficiency-weighted combination of these $dN_{sel}/d\eta(\eta, M)$ distributions leads to the final multiplicity independent pseudorapidity distribution.

4.3.4 Correction for event selection efficiency

The event selection efficiency depends on the "activity" of the collision. A more active collision (in terms of hadron production) is more likely to pass the event selection criteria than a less active. The activity in this analysis is measured by the number of pixel clusters above the cluster selection cut on a given layer. Note that neither of the event selection criteria (Section 4.1) uses information from the pixel detector. Thus, an event with zero pixel activity can be still selected, and visa versa, events with notable pixel activity may fail to pass it.

The event selection efficiency for a given type of collision, for instance, for NSD events, as a function of the number of selected clusters can be calculated as $\varepsilon_{NSD}(M) = E^{MC}_{sel}(M)/E^{MC}(M)$, where E^{MC}_{sel} denotes the number of selected NSD events in the simulation, while E^{MC} denotes all simulated NSD events. Note that M stands for the number of reconstructed hits that pass the cluster selection on a given layer. The event selection efficiency is zero for $M = 0$ since a reconstructed vertex is required during event selection. Having this correction defined in bins of multiplicity (and not as an overall integrated efficiency), the analysis is not sensitive to the precise description of the observed multiplicity distribution by the event generators.

The measured (multiplicity independent) $dN/d\eta$ distribution is built up from the efficiency weighted multiplicity-dependent yields. For inelastic collisions it would

have the simple form of

$$\frac{dN}{d\eta} = \frac{\sum_{M=0}^{\infty} \frac{E_{sel}(M)}{\varepsilon_{Inel}(M)} \frac{dN_{sel}}{d\eta}(M)}{\sum_{M>0}^{\infty} \frac{E_{sel}(M)}{\varepsilon_{Inel}(M)} + E_{Inel}^{M=0}}, \tag{4.4}$$

where $E_{sel}(M)$ is the number of selected events found in data, and $E_{Inel}^{M=0}$ denotes the number of inelastic events with zero multiplicity. This latter number is computed based on simulation events. I simply take the ratio of inelastic events with zero and non-zero multiplicity from simulation, $R^{MC} = E_{M=0}^{MC}/E_{M>0}^{MC}$, and I multiply it with selection-efficiency-weighted number of observed events in data with non-zero multiplicity: $E_{Inel}^{M=0} = R^{MC} \sum_{M>0}^{\infty} E_{sel}(M)/\varepsilon_{Inel}(M)$. The numerator in the above formula is not affected by $M = 0$ events, since $dN_{sel}/d\eta$(M=0)=0.

However, for NSD events Eq. 4.4 becomes more complicated. The reason is that contribution of SD events passing the event selection are to be removed both from the numerator and from the denominator. Hence, the equation is needed to be modified:

$$\frac{dN}{d\eta} = \frac{\sum_{M=0}^{\infty} \frac{E_{sel}(M)}{\varepsilon_{NSD}(M)} \frac{dN_{sel}}{d\eta}(M) - \frac{dN_{sel,SD}}{d\eta} \frac{E_{sel,SD}}{\varepsilon_{NSD}}}{\sum_{M>0}^{\infty} \frac{E_{sel}(M)}{\varepsilon_{NSD}(M)} + E_{NSD}^{M=0} - \frac{E_{sel,SD}}{\varepsilon_{NSD}}}, \tag{4.5}$$

where $dN_{sel,SD}/d\eta$ is the average multiplicity density of SD events passing the event selection criteria, $E_{sel,SD}$ is the number of SD events passing the event selection, and ε_{NSD} is the integrated event selection efficiency for NSD events.

In the hypothetical case of no SD events passing the event selections, $E_{sel,SD} = 0$, the above formula becomes very similar to that of Eq. 4.4. If the trigger accepts SD events, but they do not produce hits in the tracker, $dN_{sel,SD}/d\eta = 0$, only the number of accepted events is to be corrected. The reality is that the multiplicity of accepted SD events are not negligible, and thus Eq. 4.5 is to be used in its full form.

4.3.5 Comparison of measured and simulated quantities

I performed various comparison checks to ensure that the simulation describes the properties of analysis objects I see in the data. In this section I only show the most relevant quantities, such as the distribution of the pixel cluster size, the distribution of the pixel cluster charge and pixel multiplicity. The figures will be

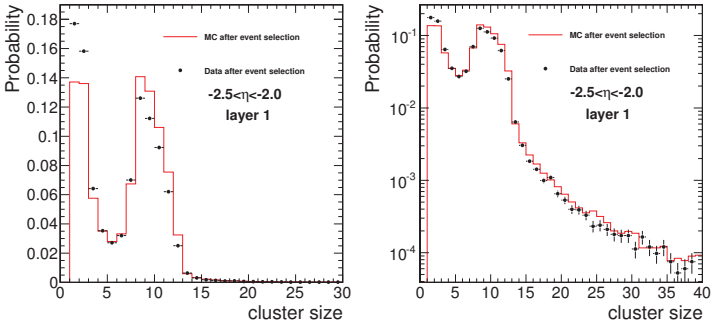


Figure 4.9: Pixel cluster size distributions in the local y direction (along the beamline) in the innermost layer after the event selection described in Section 4.1. Data are shown with black dots, the solid line shows the expectations based on simulation at 7 TeV pp collisions.

restricted to the 7 TeV data and simulation, and to a single layer. The comparison plots for the lower energy data, and for all layers are collected in Appendix B.

Distribution of cluster size The cluster size distribution in the first layer in $-2.5 < \eta < -2$ can be seen in Fig. 4.9. This distribution is a slice of the cluster size as a function η distribution, shown in Fig. 4.6 in η. The peak at low cluster sizes comes from background hits while the second peak is due to primary particles. The main features of the distribution is well described by the simulation but it underpredicts the number of clusters with low cluster sizes. As mentioned in Section 4.3.1, the cluster selection cut is defined to minimize the number of background hits above and primary hits below it. It happens that the cut defined in this way sits at the local minimum between the two peaks of the cluster size distribution. The position of the peaks as well as the local minimum is the same in data and simulation. Furthermore, the "signal" (the second peak) in the data is well described by the simulation.

The excess in data at low cluster sizes might have various sources:

- The material budget of the detector in the simulation could be wrong: in

reality particles have to travel through more material than in the simulation, thus they interact more frequently with the detector material producing more secondaries which are capable of producing small clusters even at large η.

The material budget can be probed with photon conversions and nuclear interactions in the material of the detector. These provide a detailed radiography of the material inside the CMS Tracker up to a radius of about 60 cm from the beam pipe. A good overall agreement, at a level of $\sim 10\%$, is observed between the distribution of the material in the real and in the simulated Tracker detector, with only a limited number of localized discrepancies [50]. This fact rules out the material budget based explanation of the excess at small clusters.

- In data the fraction of "broken" (split) pixel clusters is larger than in the simulation. To understand the origin of broken clusters we need to recall what a pixel cluster is. A cluster is a set of neighbouring pixel units which got above the readout threshold due to the deposited charge of a particle which traversed the sensor. If one of the pixel units fails to detect[22] the necessary charge to be read out, the cluster might split into two parts creating small clusters. The splitting of clusters (above the expectation based on simulation) would also affect the final results of the analysis. Hence, it is important to estimate the relative abundance of broken clusters in data and in simulation.

  Two clusters of a split cluster are close to each other. Thus, a useful handle of the split clusters is the relative abundance of cluster pairs, which are separated by no more than a couple of pixel units. Following this idea, I computed the minimal distance of each cluster from the rest of the clusters (Fig. 4.10). This is done after all the event selection cuts are imposed, both in data and in simulation. The minimal distance is calculated with respect to all pixel units of the clusters. Thus, the different shapes of clusters are properly accounted for. The pixel units in the barrel has a surface area of $150 \times 100\ \mu m^2$. I estimate the number of broken clusters as if every cluster closer to each other than 600 μm would be a fragment from a split cluster. This definition is a

[22] Due to wrong pixel readout timing, malfunctioning pixel module, etc.

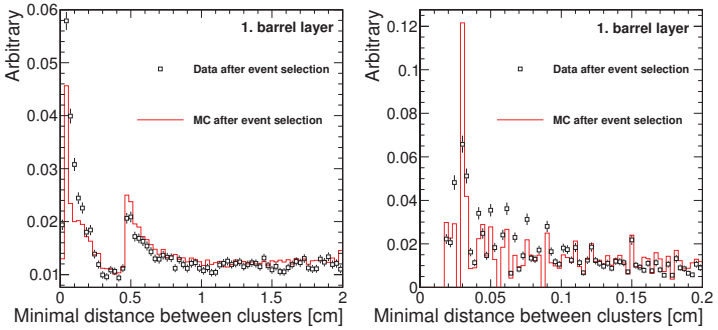


Figure 4.10: The distribution of minimal distance between pixel clusters in the first pixel barrel layer. The red histogram shows the simulation expectation, the black symbols are the measured values.

Energy [TeV]	layer 1 [%]		layer 2 [%]		layer 3 [%]	
	MC	Data	MC	Data	MC	Data
0.9	0.9	1.5	0.6	1.2	0.5	1.0
2.36	0.9	1.5	0.6	1.1	0.5	1.0
7	0.8	1.5	0.7	1.1	0.4	1.0

Table 4.2: Fraction of split clusters in data and in simulation at 900 GeV, 2.36 and 7 TeV.

purely technical one, the absolute amount of split clusters resulting from this study is a conservative upper limit on the "real" frequency of splitting. From the point of view of the analysis, only the observed difference of this frequency in data and simulation is important, not the absolute amount. The fraction of such broken clusters in simulation and in data at 900 GeV, 2.36 TeV and 7 TeV are given in Table 4.2. One can see twice as many "split clusters" in data than in simulation, but the overall number of broken clusters in data is still small, thus the difference is small, too. Thus, I don't correct for them in the analysis but incorporate this effect into the systematic error of the measurement.

- Low-p_{T} particles ($p_{\mathrm{T}} < 1$ GeV/c) are produced in greater relative abundance in real collisions than in the simulations. Since a part of the low-p_{T} particles are loopers, they are capable of causing small clusters even at large η. The

measurement of transverse momentum provided by the tracking method (Section 4.5) indeed shows that there is clear excess of low-p_{T} particles in data compared to simulation (see in Section 5).

Since the deviation is concentrated in the first two multiplicity bins, which are cut away by the cluster selection cut, no extra correction is applied in any of the η bins but in the most central ones, where the χ correction is increased according to the observed excess at larger η.

Distribution of cluster charge In this paragraph the angle corrected pixel charge distributions (charge/cosh(η)) are shown and discussed. In the above sentence "angle corrected" has the following meaning. The cluster charge is proportional to the length of the particle trajectory in the sensitive detector layer, which created the cluster. Thus, the distribution of the pure cluster charge depends on the spatial orientation of produced particles. To remove this dependence, the cluster charge is to be normalized by the length of trajectory within the pixel layer. Since in this analysis I do not have access to the trajectories of particles, the length of the trajectory is estimated from the cluster position assuming that all particles are primaries and propagate along a straight line: trajectory length $\sim \cosh(\eta)$.

The angle corrected pixel charge distributions in data and in simulation for all pixel clusters are shown in Fig. 4.11. The peak at low corrected cluster charge consists of background clusters (the trajectory length assigned to such clusters is computed as if they were primaries, which is an overestimation), while the second peak is due to primary particles. The plot shows the same feature already observed in the cluster size distributions (Fig. 4.9): excess of background clusters. Once the cluster selection cut is applied, the agreement becomes much better (Fig. 4.12), the distributions show a single Landau peak meaning that the background is largely suppressed.

Distribution of cluster size along the beamline The cluster length distributions along the beamline (in the *global z* or in the *local y* direction) in the first pixel barrel layer ($|\eta| \leq 2.5$) is shown in Fig. 4.13. A good general agreement is found,

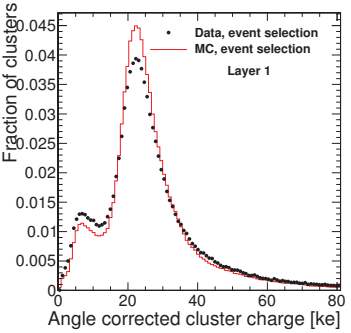


Figure 4.11: Angle corrected cluster charge distribution in the first pixel barrel layer in simulation and in data after event selection.

although, the already observed excess is visible at low cluster length. Note that here all clusters in the pixel detector is used, not just those within a η interval, as for Fig. 4.9. On the right panel, a single bin at a cluster size of 48 shows a clear deviation from the simulation. This single bin has negligible effect on the analysis, but nevertheless the source of this deviation was investigated and was tracked down to a problem in the local cluster reconstruction algorithm.

Pixel cluster multiplicity The multiplicity of pixel clusters is the quantity I measure in the analysis. Thus, on the one hand, no precise description is needed in this observable by the simulation. On the other hand, since the simulation-based corrections are computed in bins of multiplicity, it is important to have the multiplicity range in data covered by the simulation. This requirement had a practical implication during data analysis. The PYTHIA D6T [51] tune (the standard tune of CMS), which I used for the analysis of the 0.9 and 2.36 TeV data was replaced by the PYTHIA ATLAS tune for the 7 TeV data. The reason was that the high multiplicity tail observed in data was not present in D6T, which made the computation of correction factors impossible. The comparisons between the measured and simulation-based event-by-event pixel multiplicity distributions at 7 TeV can be seen in Fig 4.14. The simulation does not reproduce the observed multiplicity,

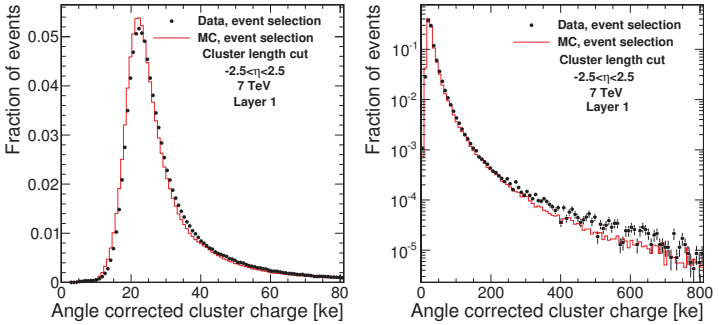


Figure 4.12: Cluster charge distributions in the first barrel layer of the pixel detector after event and cluster selection. Data are shown with black dots, the solid line shows the simulation expectation.

but provides the necessary multiplicity coverage.

Clusters, which are not on reconstructed tracks One way to study the properties of background clusters is via tracking (see Section 4.5). The idea is that one reconstructs the trajectories of primary particles, and then selects those clusters, which are not associated to any of the trajectories. Most of such clusters are background clusters from the point of view of the cluster counting analysis[23]. This can be done in data and also in simulation; and the properties of the two sets of background clusters can be compared. As a result of this study, I found that the properties of such background clusters are well described by the simulation. In Fig. 4.15 the angle corrected cluster charge distributions of such clusters are shown for data and simulation. The different magnitudes of the distributions show that in data more clusters are produced than in the simulation, but the shapes of the distributions are very similar. Indeed, the difference is only a factor of ~ 1.404. Assuming that a simple naive scaling of the primary charged particle yield, as it is present in the simulation, is possible with this factor, one gets $1.404 \times dN/d\eta\ (sim) \approx 5.8$, which is very close to the real measured value (see Section 5).

[23]The pixel detector is completely noise free: there were only a few noisy clusters but they were masked in the reconstruction software of CMS.

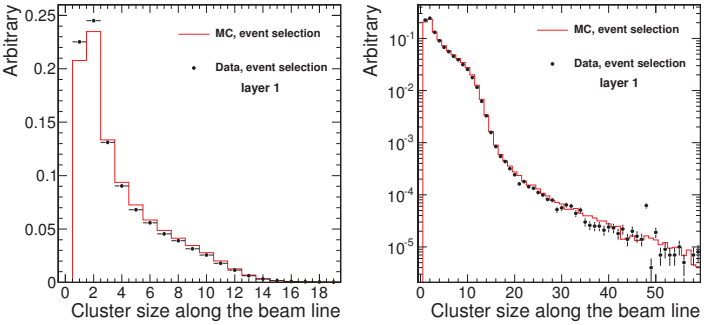


Figure 4.13: Cluster length distributions in the first barrel layer of the pixel detector along the beamline after event selection. Data are shown with black dots, the solid line shows the simulation expectation.

4.4 The tracklet method

As mentioned in the introduction of this section, the tracklet method can be regarded as a coincidence version of pixel counting, and was developed by my colleagues at MIT. Since this analysis was (together with the pixel counting and tracking methods) performed with the first collision data, I devoted a short subsection to briefly introduce it.

Pixel tracklets are pairs of pixel clusters on any two of the three barrel layers. The difference in the angular positions of the two clusters with respect to the PV, $\Delta\eta$ and $\Delta\phi$, are calculated for each tracklet. If two tracklets share a hit, the tracklet with the larger $\Delta\eta$ is discarded. Tracklets, which are associated to primary particles have a sharp peak in $\Delta\eta$ around 0, and two peaks around $\Delta\phi = 0$ (the two peaks appear due to the magnetic field). Tracklets also contain "fake" cluster pairs due to the large combinatorial possibilities, and tracklets associated to secondary particles. The combinatorial background tracklets are suppressed by restricting the allowed $\Delta\eta$ and $\Delta\phi$ of tracklets. The remaining combinatorial tracklets exhibit an extended tail in both the $\Delta\eta$ and $\Delta\phi$ distributions, which is flat in case of $\Delta\phi$. The combinatorial background can be subtracted based on this flatness: sidebands are defined on both sides of the dual signal peak to determine the background at the position of

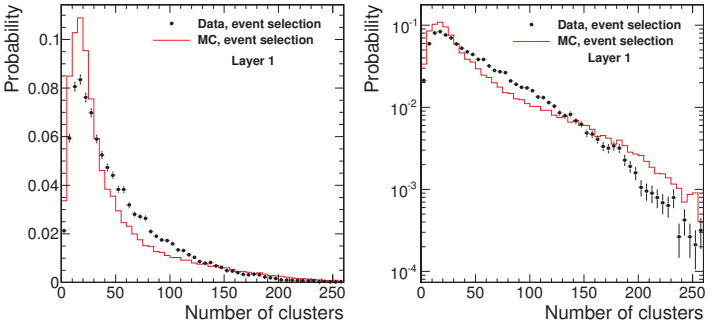


Figure 4.14: Event-by-event cluster number multiplicity in the first barrel layer of the pixel detector along the beamline at 7 TeV, after event selection. Data are shown with black dots, the solid line shows the PYTHIA ATLAS simulation.

the peak. The contribution from secondary particles, together with reconstruction efficiency, and geometrical acceptance needs to be accounted for to determine the number of primary charged hadrons. These correction factors were calculated using the simulations for background-subtracted tracklets in bins of z position of the PV, pseudorapidity, and tracklet multiplicity.

4.4.1 Pixel cluster efficiency

Since all of the measurements presented in this chapter use pixel hits, the measurement of pixel cluster reconstruction efficiency (this is the probability that a cluster indeed emerges due to a passing charged particle) is of vital importance.

The pixel cluster efficiency can be measured by a data driven method of inspecting hits on a layer extrapolated (or interpolated) from the tracklets of the other two layers. The important quantity from the point of view of the measurements, is not the absolute pixel cluster efficiency, but the *double ratio* of the efficiencies from data and simulation. This is because the hit efficiency correction, due to the nature of the analyses, is folded into the corrections taken from the simulation.

The main philosophy is to "tag" a pixel tracklet and "probe" if there is a hit at the expected position on the other layer. One can inspect the efficiency of 1^{st}-layer

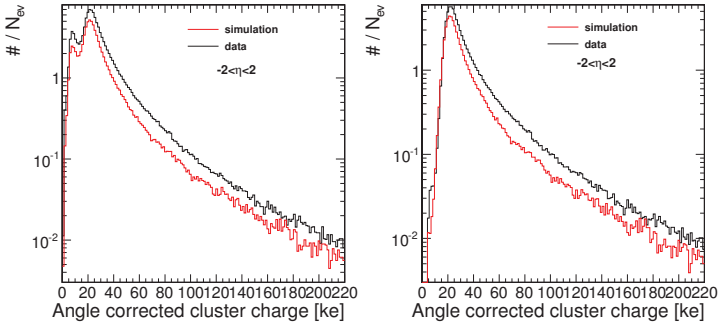


Figure 4.15: Distribution of angle corrected cluster charge of clusters which are not associated to tracks; *left*: all clusters, *right*: clusters above the cluster size cut.

Layer	Efficiency in data [%]	Efficiency in MC [%]	Double Ratio×100
1	91.9 ± 0.1	92.6 ± 0.1	99.2 ± 0.1
2	98.7 ± 0.2	97.4 ± 0.1	101.3 ± 0.2
3	95.7 ± 0.1	95.6 ± 0.1	100.1 ± 0.1

Table 4.3: Efficiencies and double ratios in data and simulation. Only the statistical uncertainties are shown.

hit reconstruction from the 2-3 tracklet (tracklet from 2^{nd} and 3^{rd} barrel layers), as well as 2^{nd}-layer hits from 1-3 tracklets, and 3^{rd}-layer hits from 1-2 tracklets.

The found double ratios are within 1.3% from being unity (the results are summarized in Table 4.3). This is rather small and thus no extra corrections were introduced to account for it but it was built in the systematic uncertainties of the measurements.

The very same idea applied here can be used to "map" the sensitive surface of the pixel detector and compare it with those in the simulation (Appendix B.3). The results of this study was used during the analysis of data to take the effect of the inactive parts of the detector into account.

4.5 The tracking method

The method described here, the so-called minimum bias tracking, was developed to allows us to reconstruct particles down to $\sim$ 100 MeV/c. The standard CMS tracking method was only able to reconstruct tracks above $\sim$ 500 MeV/c) with high *efficiency* and low *fake* track rate (to be discussed below). As described in Section 1.3, the smaller the minimal p_{T} of the method is, the smaller the systematic uncertainties of the $dN_{\mathrm{ch}}/d\eta$ values (integrated over p_{T}) are. Since the most probable p_{T} is around a few hundred MeV/c, pushing down the minimal p_{T} from 500 to around 100 MeV/c is very important.

4.5.1 Steps of track reconstruction

During track reconstruction all the barrel and endcap layers of the Pixel and Strip Detector are used. The algorithm starts with the assumption that the position of the beamspot is already known. In general this is true and the beamspot information is provided by the Tracking Group of CMS using the standard tracking algorithm.

The tracking algorithm builds the tracks starting from the layers closest to the interaction and propagates them outwards. First, so-called *seeds* are produced using the pixels, then proto-tracks are grown into the direction determined by the seeds, into the Strip Detector, where further hits are added to the proto-tracks. The algorithm terminates when the proto-track reaches the outmost tracker layer, or if it fails to propagate it any further. The details of the minimum bias tracking algorithm are as follows.

In the first step pixel cluster pairs are formed.

Then these pairs are given to the so-called triplet building algorithm. This algorithm takes the pairs and tries to find a third hit in the next detector unit (proceeding outwards) by performing the following exact geometrical calculations. A cylinder ("cylinder of origin") is defined around the beamline with a radius of 0.2 cm and a length of $3\sigma_z$ of the beamspot in both directions, where σ_z is the width of the Gaussian profile of the interaction region along the beamline. In the small volume of the pixel detector the magnetic field is practically constant, the

charged particles propagate on helices. The projection of a helix or a cylinder onto the transverse plane is a circle. With the given cylinder of origin, and the point pair P_1 and P_2 of the pixel cluster pair, the set of allowed trajectories originating from the cylinder and traversing the two points can be constructed. The set of possible trajectories defines the search window for the third hit of the track in the next pixel layer as follows. The possible position of the third hit in the transverse direction is enclosed by a pair of so-called limiting circles, projection of extreme trajectories (left panel of Fig. 4.16). The standard method in this step uses a simple straight line prediction, connecting the beamline with the outer hit, to define the search window for the third hit. For low-p_{T} particles, the straight line approximation does not hold. However, the exact calculation using limiting circles still works even at low-p_{T}. For high-p_{T}, the difference between the two methods vanishes. The combination of the point pair P_1 and P_2 with a candidate third hit, P_3, defines unambiguously a circle in the transverse plane. The helix has a linear relationship between the azimuthal angle with respect to the center of the circle and the z coordinate. This way, with help of the positions of P_1 and P_2, the z coordinate of P_3 is determined, too. The procedure outlined above creates pixel triplets out of doublets.

Since at low-p_{T} the allowed region of the third hit can be relatively large, *fake* triplets (triplets not associated to any particle) might be created. To remove these undesired triplets, the information present in the shape of the pixel clusters can be used: the shape of the cluster can be compared with the local direction of the trajectory. Triplets with incompatible cluster sizes are removed.

As the next step, the interaction vertex is reconstructed as described in Section 4.2 using triplets. This results in a proto-vertex; only used as an additional constraint during the next steps of track reconstruction.

Once the vertex (or vertices) is found an other round of triplet finding starts. All the previous triplets are decomposed, the triplet building algorithm starts from the beginning, but this time it uses the position of the interaction point. The vertex information is exploited to restrict the cylinder of origin: the radius of the cylinder remains 0.2 cm, but the length is reduced to $3\sigma_z^{vtx}$ in both directions. The minimal transverse momentum of triplets, $p_{\mathrm{T}}^{\mathrm{min}}$, is set to 75 MeV/$c$. All triplets with smaller

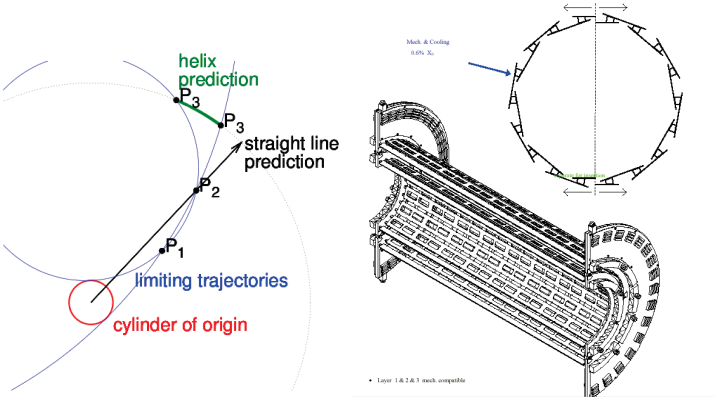


Figure 4.16: *Left*: Schematic comparison of the standard straight line prediction (black line) and the new helix prediction (green arc) for finding the third hit. Limiting trajectories (blue circles) that touch the cylinder of origin (red circle) cut out an arc from the barrel layer of the third hit candidates. *Right*: Schematic view of the pixel detector. A layer is composed of two *ladders* which overlap at certain points of the detector. Thus particles traversing through the layer might leave more than one hit in it.

p_{T} are removed. Finally, the cluster shape based triplet filter is applied again.

The resulting triplets always consist of three hits, but several of them might be related to a single track of a single particle. This can happen due to the overlapping nature of pixel units (right panel of Fig. 4.16). Given the overlaps, it is possible that more than 3 hits are left in the Pixel Detector by a traversing particle. As a consequence of this, more than one triplet can be created out of these multiple hits. Thus, a cleaning and merging procedure of the triplets was needed to be developed to resolve the problem of such *multiple counting*. This procedure work as follows. If two triplets share at least half of the hits of the shorter one (triplets start as cluster triads, but they can grow due to the iterative procedure described here) and a) the remaining hits are on different detector units, then they are merged by adding all the hits of one of the triplets to the other, the other one is deleted; b) the remaining hits are on the same detector unit, then the one with smaller χ^2 is kept, the other is removed. If the number of shared hits is smaller than the number of clusters of the

shorter triplet, but larger than 1, the longer triplet is chosen or in case of equally long triplets, the one with smaller χ^2. If only 1 hit is shared, both triplets are kept (this happens, for instance, in case of photon conversions).

Once these *extended triplet seeds* of the to-be-built tracks are formed, they are propagated through the whole Tracker Detector picking clusters from the Strip Detector.

The propagation of tracks is done by the so-called Kalman filter [52]. This approach builds upon the knowledge of detector material and local hit position resolution, using the known physics of multiple scattering and energy loss. It is widely used in present particle physics experiments for tracking and provides a coherent framework to handle known physical effects and measurement uncertainties. It is equivalent to a global linear least-squares fit which takes into account all correlations coming from process noise. It is the optimum solution, since it minimizes the mean square estimation error. The cluster shape filter is always present during the propagation of seeds, although it has smaller cleaning power in this step than it has during triplet building, because the strip clusters are one dimensional as opposed to the two dimensional pixel clusters.

The resulting tracks (composed of pixel and strip hits) are saved and their clusters are removed from further considerations. Such tracks compose the first set of the final track collection. Due to the closed cylinder of origin (radius: 0.2 cm), they mainly consist of tracks due to primary particles.

Given that in the previous step most of the clusters are removed, the combinatorial possibilities of formable tracks significantly decreased. The cylinder of origin can be opened now without significantly increasing the fake rate: its radius is increased to 3.5 cm (out to the beampipe). The vertex requirement is also dropped during this round of formation of triplets. The tracks are then reconstructed as described above. The main purpose of this track reconstruction step is to find the remaining primary particles and to reconstruct secondaries. The tracks from this reconstruction step are merged with the tracks of the previous step, resulting in a extended track collection. Then the clusters used to build tracks are removed, just as in the previous step.

There are still untracked charged particles left after all the steps above. The reason is that particles might traverse inactive detector material leaving no hit. Such particles cannot be reconstructed with triplet seeds, since they do not have three hits. They can only be tracked with pair seeding, which is the final step of the tracking procedure.

The radius of the cylinder of origin is decreased to 0.4 cm and the vertex constraint is added back again to the seeding step, which here only passes through the pixel pairs created from all layer combinations (except pure endcap pairs) as input to the next module of the tracking algorithm. If two pixel pairs have a common cluster, one of them (randomly selected) is decomposed. The pair seed is given to the Kalman filter which propagates it through the Tracker Detector. A resulting track is only saved if it has at least 3 hits; then they are merged with the previously found tracks. Thus, the final collection of tracks are complete.

Although the track collection is complete, the tracking algorithm has two more steps. The positions of vertices are updated using the tracks from the final track collection as inputs for the reconstruction algorithm, and the primary tracks are selected from the track collection based on the compatibility of the trajectories with being created on the interaction vertex.

Primary tracks are selected based on their compatibility with the beamspot in the transverse, and the primary interaction vertex in the longitudinal direction. In the transverse plane the tracks need to be within the $5\sigma_{\mathrm{T}}$ vicinity of the beamspot, where σ_T consists of the uncertainty on the track and beam positions, $\sigma_T^2 = \sigma_{\mathrm{T,beam}}^2 + \sigma_{\mathrm{T,track}}^2$, $\sigma_{\mathrm{T,beam}}^2 = (\textit{beam width in } x \times \textit{beam width in } y)$; the distance must always be smaller than 0.2 cm. In the longitudinal direction the tracks must be closer to the vertex than $5\sigma_L$, where $\sigma_{\mathrm{L}}^2 = \sigma_{\mathrm{L,track}}^2 + (\cosh\eta)^2\sigma_{\mathrm{T,beam}}^2$.

The steps of track reconstruction, as described above, are summarized in Table 4.4.

4.5.2 Corrections

For a physical result, the tracks in the track collection are to be corrected for various inefficiencies. There are track level (for instance, reconstruction efficiency)

Table 4.4: Summary of tracking steps. For the details see Section 4.5.1.

Step	Seed	Use vtx?	r_{orig}	Remark
0	triplets	no	0.2 cm	pixel tracks for vertexing
$\text{vtx}_{triplets}$				vertexing with triplets
1	triplets	yes	0.2 cm	primaries from triplets
2	triplets	no	3.5 cm	secondaries from triplets
3	pairs	yes	0.4 cm	primaries from pairs
vtx_{tracks}				vertexing with tracks
primaries				selection of primaries

and event level (for instance, event selection efficiency) corrections.

In order to be able to evaluate the performance (and so the efficiency) of the tracking algorithm, a few notations are needed to be introduced:

- A simulated charged particle is *accepted* if it leaves enough hits to be reconstructed: at least two hits in the pixel detector.
- The *acceptance* (or *geometrical acceptance*) is the fraction of the accepted primary simulated charged particles.
- A track is *associated* to a simulated charged particle if more than half of its hits are shared with a simulated charged particle.
- A track is *fake* if it cannot be associated to any simulated charged particle.
- A simulated charged particle is *counted n times* if there are n tracks associated to it. A simulated particle is counted multiple times, if $n > 1$.
- *Efficiency* (or *algorithmic efficiency*) is the fraction of single counted accepted primary simulated charged particles.
- *Fake rate* is the fraction of fake primary tracks.
- *Multiple counting* is the fraction of multiple counted simulated charged particles.

In this section only the figures made using simulation at 7 TeV are shown: the discussed quantities have very little collision energy dependence, since they mostly depend on the properties of the given particle created in the collision, not on the way the particle was created.

Acceptance Studies with simulated events show that the acceptance is flat both in η and p_{T}. The acceptance extends down to p_{T} of about 0.05, 0.1 and 0.2 GeV/c for pions, kaons, and protons, respectively (Fig. 4.17). This visible mass dependence is due the different amount of energy loss these particles suffer traversing material at a fixed p_{T} (heavier particles lose more energy at small momenta). The edges of the acceptance as a function of η is set by the spatial coverage of the tracker system as well as the the smearing of the vertex position along z. The acceptance is above 95% in the $|\eta| < 2$ and $p_{\mathrm{T}} > 0.2$ GeV/c regions.

The results for unidentified charged hadrons practically coincide with the pion values, since most of the produced charged particles are pions.

Efficiency The algorithmic efficiency (Fig. 4.18) has similar p_{T} and mass dependences as the acceptance. The efficiency differs for pions, kaons, and protons due to the different amount of multiple scattering they suffer at a certain p_{T} (the multiple scattering depends on β, not on p_{T}). The efficiency is around 80–85% for midrapidity. The practical minimal p_{T} of track reconstruction is around 0.1, 0.2 and 0.3 GeV/c for pions, kaons, and protons, respectively.

Fake rate The fake rate is kept below the percent level (Fig 4.19) by exploiting the geometrical shape of the pixel clusters, as already discussed. Note that in the case of endcap clusters the filtering does not have a strong differentiating power, since the local trajectory direction is close to perpendicular. Endcap clusters are only contained in forward tracks, thus the maxima of the fake rate are at forward η.

Multiple counting The probability of multiple counting is very small due to careful triplet cleaner and merger. It is well below the percent level (Fig 4.20).

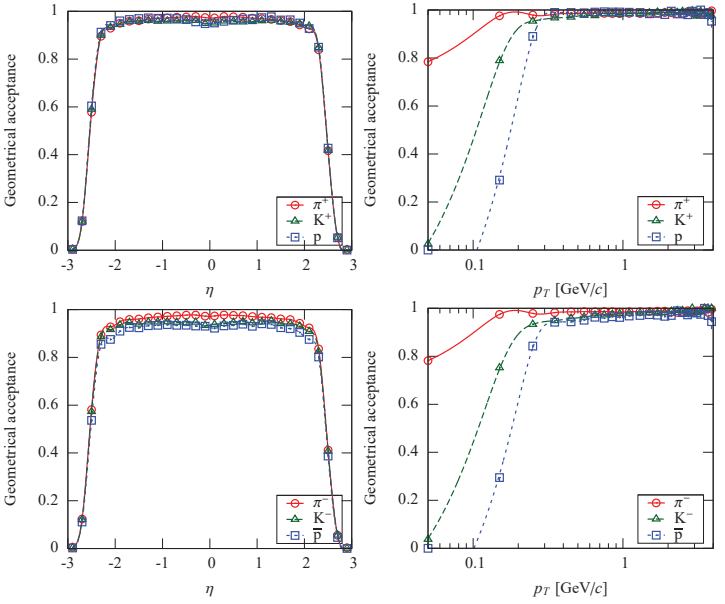


Figure 4.17: Geometrical acceptance as a function of η (*left*) and p_{T} (*right*), for positive (*top*) and negative (*bottom*) particles. Values are given separately for pions (red circles), kaons (green rectangles), and protons (blue boxes). The curves are drawn to guide the eye.

Some characteristics of the tracking are expected to depend slightly on the multiplicity of the event. The acceptance is independent of the multiplicity, since it only depends on detector geometry. In proton-proton collisions, even with a large number of pile-up events, the occupancy of the tracker remains small, thus the algorithmic efficiency is insensitive to multiplicity, too. However, the multiple counting and fake rate might have a slight multiplicity dependence. The multiple counting does not show any sensitivity to multiplicity (left panel of Fig. 4.21), but the fake rate does (right panel of Fig 4.21). It is roughly proportional to the number of reconstructed tracks. More reconstructed tracks mean more particles (primaries and secondaries) leaving hits in the silicon detectors. This facilitates the creation of

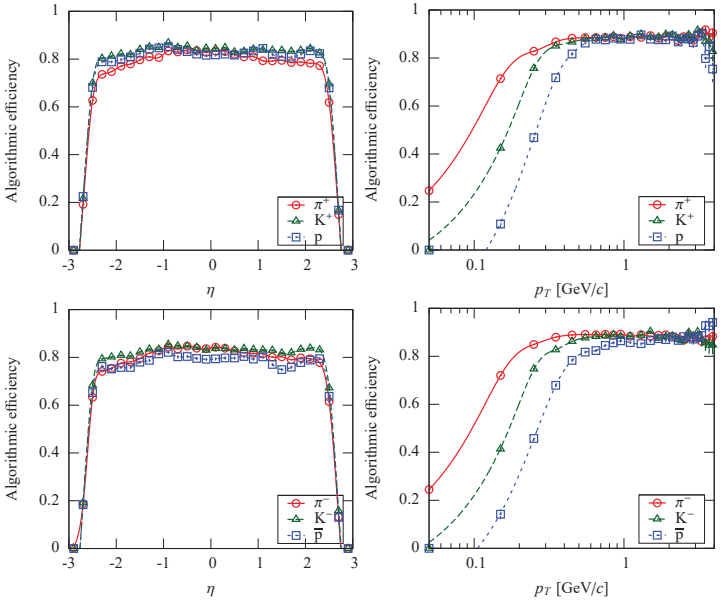


Figure 4.18: Algorithmic efficiency as a function of η (*left*) and p_{T} (*right*), for positive (*top*) and negative (*bottom*) particles. Values are given separately for pions (red circles), kaons (green rectangles), and protons (blue boxes). The curves are drawn to guide the eye.

fake hit triplets in the pixel detector. Thanks to the efficient track and trajectory filtering and cleaning procedures, as well as the use of the predetermined primary vertex, the dependence is only linear instead of a higher order.

The charged hadron spectra has to be corrected for particles of non-primary origin (secondaries) passing the primary selection. The main source of these secondary particles is the *feed-down* from weakly decaying particles, mostly K^0_s, $\Lambda/\bar{\Lambda}$ and $\Sigma^+/\bar{\Sigma}^-$ (see Table 4.5). As these particles decay into charged pions and protons (and not into kaons), only a negligible correction is expected for charged kaons.

Pions and protons coming from decays happening close to the primary vertex may look as if they were primaries. This is especially true in the low-p_{T} region, since

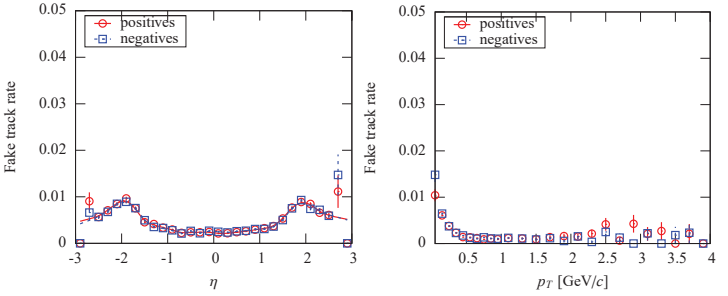


Figure 4.19: Fake track rate as a function of η (*left*) and p_{T} (*right*), for positive and negative particles. The curves are drawn to guide the eye.

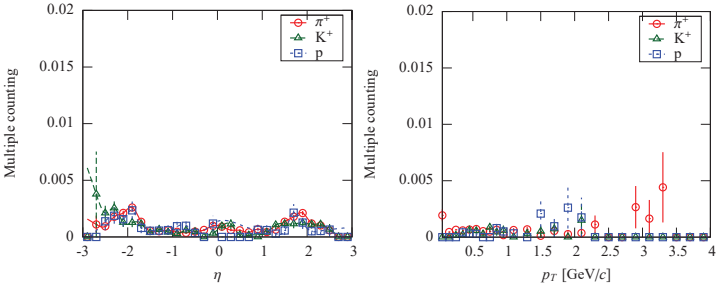


Figure 4.20: Multiple counting as a function of η (*left*) and p_{T} (*right*), for positive and negative particles. The curves are drawn to guide the eye.

the significant multiple scattering and energy loss effects wash out the differences between primary and secondary particles. Note that similar corrections are expected for positive and negative hadrons, because the production of resonances at LHC energies are closely charge independent. Secondary particles from inelastic collisions of primary particles with detector and other supporting material are also important contributors to this correction. In fact, for kaons, this is the only relevant source of non-primaries among the selected tracks.

The fraction of secondary particles in the sample of reconstructed primary particles, obtained from the PYTHIA generator, is shown in Fig. 4.22. While the correction is around 2% for pions, it is much larger and can go up to 15% for protons

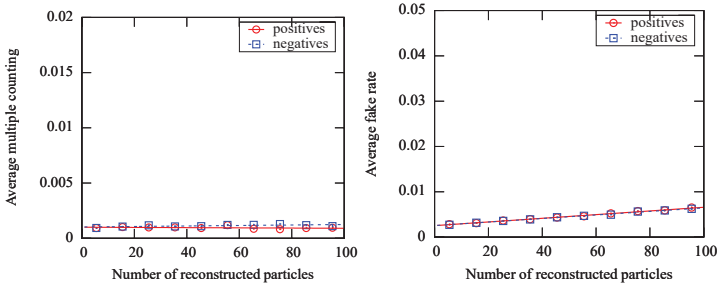


Figure 4.21: Multiplicity dependence of multiple counting (*left*) and fake rate (*right*) for positive and negative particles. The curves are linear fits.

Resonance	Decay	Branching	$c\tau$ [cm]
K^0_s	$\pi^+\pi^-$	69.2%	2.68
$\Lambda/\bar{\Lambda}$	p π^- / $\bar{p}$ π^+	63.9%	7.89
$\Sigma^+/\bar{\Sigma}^-$	p π^0 / $\bar{p}$ π^0	51.6%	2.40

Table 4.5: The most abundantly produced weakly decaying resonances and their relevant decay channels, branching ratios and average decay length values.

at $p_{\mathrm{T}} \approx 0.1$ GeV/c. The observed difference in the correction between pions and protons is expected because the daughter p or $\bar{p}$ will take most of the momentum of the primary Λ or $\bar{\Lambda}$, therefore its trajectory will more likely extrapolate (wrongly) to the primary interaction point than a pion coming from a K^0_s decay.

The reconstructed tracks are binned in two-dimensional histograms $\Delta N(\eta, p_{\mathrm{T}})$. The multiplicity of charged particles, for a given bin, can be computed using the number of tracks and the above-mentioned corrections in the following way:

$$\Delta N_{\mathrm{corrected}} = \frac{(1 - \mathrm{FakeRate})(1 - \mathrm{NonPrimaries})}{\mathrm{Acceptance} \cdot \mathrm{Efficiency} \cdot (1 - \mathrm{MultipleCounting})} \Delta N_{\mathrm{measured}} \qquad (4.6)$$

Note that here the charged particle multiplicity is not yet corrected for the event selection efficiency. That is done by computing the event selection efficiency $\varepsilon_{Eff} = E^{MC}_{sel}/E^{MC}_{NSD}$ as a function of track multiplicity. This ε_{Eff} is then applied as an event weight when the summation over multiplicity is done.

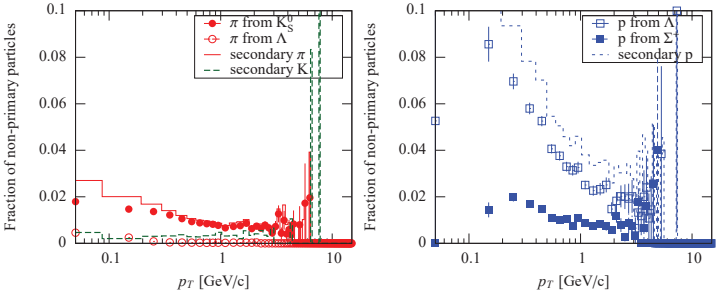


Figure 4.22: Fraction of non-primary particles in the sample of reconstructed primary particles for pions, kaons, and protons as a function of p_{T}. Sources of non-primaries are feed-down from weakly decaying resonances and inelastic collisions of primary particles with detector and supporting material.

4.6 Systematic uncertainties

In this section the various systematic uncertainties affecting the three measurement methods are discussed. The tests done to examine the sensitivity of the applied analysis methods to various systematic uncertainty sources will already refer to the analysis of real data. The results will be presented in the next section, Section 5. The studies presented here refer to the 0.9 and 2.36 TeV data, but were also performed at 7 TeV. First, the uncertainties are listed one-by-one followed by a short discussion on each of them, then the summary of the estimated uncertainties are shown as a table, Table 4.6.

- **Event selection**. The sensitivity of the final results to the event selection efficiency can be tested using various triggering and event selection strategies. I varied the final strategy described in Section 4.1, by omitting completely the offline HF cut, and introducing a BSC-coincidence-based selection. The results were found to differ by less than 3%. The BSC MIP (minimum ionizing particle) efficiency was also varied by 5%, causing about 1% change in the efficiency of the used BSC triggers, changing the $dN/d\eta$ result only by 0.5%. Based on such sensitivity checks we assigned a systematic uncertainty of 3%.

- **Detector acceptance**. The quality of the description of the sensitive surface of the pixel detector by the simulation might be affected by unknown inactive pixel modules in the data. Hence we developed a method to map the pixel detector, which was briefly mentioned in Section 4.4.1. We found only a single dead pixel readout module compared to simulation. In the analyses I have taken this module into account by rejecting the simulated clusters in that detector area.

  The position of the pixel barrel modules are a bit different in simulation than in data, thus the edges of the η acceptance (extremities of the pixel barrel layers in z) are not used in the analyses. The (η, ϕ) acceptance was plotted from data and simulation, and their ratio was smeared with the measured vertex z distribution. From this study, the acceptance uncertainty was estimated to be 1%.

- **Pixel cluster efficiency**. The probability that a passing charged particle leaves a hit in a pixel layer is larger than 99%, as measured from data (Section 4.4.1). This result is also consistent with the matching distributions of the number of pixel hits on tracks distribution, which has shown a very good agreement between data and simulation (Fig. 4.23). The pixel hit efficiency affects tracklets the most, since in the tracklet reconstruction two layers are required; an inefficiency of 0.5% per pixel layer leads to an inefficiency of 1% for tracklet finding. Tracks are not much affected thanks to the pair seeding step of track reconstruction[24]. The cluster counting analysis is affected by the amount of uncertainty itself.

- **Pixel cluster splitting**. The relative fraction of split clusters can be estimated using a simple cluster-distance-based method (Section 4.3.5). I did not observe a larger difference between the fraction of split clusters in data and simulation than 0.5-0.7%. I did not apply any correction for split clusters,

[24]Let the average hit loss per layer be p. The chance to loose a hit triplet is thus $3p^2(1-p)$. In case of pairs the same probability is $2p(1-p)$. Since the ratio of pair seeds to triplet seed is ~ 0.1, the average track loss is $0.9 \cdot 3p^2(1-p) + 0.1 \cdot 2p(1-p)$. In case of $p = 2\%$ the average track loss is only 0.5%

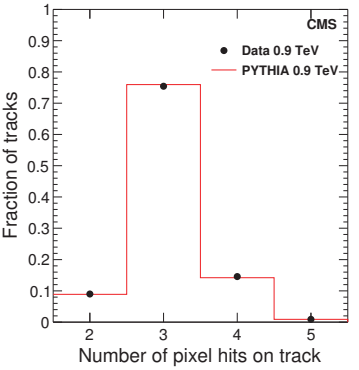


Figure 4.23: Distribution of pixel hits attached to tracks in $|\eta| < 1$ (dots) compared to simulation (histogram).

but assigned a conservative systematic uncertainty of 1% for cluster counting. The other analyses are less affected.

- **Tracklet and cluster selection**. By varying the cuts used to select clusters (cluster selection cut, see Section 4.3.1) and tracklets ($\Delta\eta$ and $\Delta\phi$ cut, see Section 4.4) the sensitivity of the analyses to the cuts applied can be estimated. The analyses were repeated using various cuts. Based on the observed changes in the final results an systematic uncertainty of 3% and 0.5% were assigned, respectively.

- **Algorithmic efficiency**. For the pixel counting analysis no high-level objects are reconstructed, one cannot speak about algorithmic efficiency. For tracklets the reconstruction efficiency depends on the shape of the combinatorial background which was tested with various simulations, resulting in a systematic uncertainty of 3%. The tracking analysis has mass-dependent corrections at low-p_{T}, thus the algorithmic efficiency depends on the particle composition. Since during these measurements no particle identification was possible, the abundance of particle species is taken from simulation. A systematic uncertainty of 2% is assigned based on changing the composition between reasonable

limits.

- **Loopers**. The contribution of loopers to the number of clusters above the cluster selection cut is about 4%. I assigned the half of this correction as the systematic uncertainty. This fraction is removed by the looper correction which is based on the amount of short clusters seen in the data (Section 4.3.2). Tracklets are less sensitive to loopers at large η due to the $\Delta\eta$ cut; we assigned a uncertainty of 1%. The tracking method is almost completely insensitive to them thanks to the selection of primary particles.

- **Secondary particles**. In data the material budget was checked and compared to that in the simulation (Section 4.3.5 and Ref. [50]), where a good agreement was found. The χ correction of cluster counting comes from simulation, and thus, it was checked using modified simulation. The uncertainty on the correction was found to be $\sim 2\%$. The other methods are less sensitive to secondaries; a systematic uncertainty of 1% is assigned for them.

- **Misalignment**.

  Pixel counting is not sensitive at all to the small possible misalignment of the tracker. The tracklet method has a sensitivity through the $\Delta\eta$ cut, which was studied by moving the reconstructed hit positions while keeping the vertex position at the same place, resulting in an uncertainty of 1%. Tracks below $\sim$ 10 GeV/c are again not sensitive to misalignment since the hit position uncertainty is dominated by multiple scattering. Still, a systematic uncertainty of 0.1% is assigned.

- **Other background hits**. Thanks to the extremely clean event selection in use (Section 4.1), none of the beam-gas events[25] survive and get into our data sample. This was tested on filled-empty bunch crossing data, where zero events survived the event selection cuts.

  The other source of possible background is the accidentally overlapping beam-gas and collision events, but these are very rare, since the beam-gas rate and

[25]Collisions between the protons in the beam and remnant (due to imperfect vacuum) molecules/atoms in the beam pipe.

the collision rate was both less than 50 Hz (in case of beam gas it is much less). Besides, tails of the cluster size distribution would appear in the data if such a contribution existed, but that is not seen compared to simulation (see Section 4.3.5).

Events with beam halo particles are vetoed in our event selection (Section 4.1.2). Thus, their contribution to the number of hits is negligible.

We assigned a conservative uncertainty of 1%, 0.2%, and 0.1% for the cluster counting, tracklet, and tracking methods, respectively.

- **Multiple counting, fake rate**. Multiple counting and fake rate was shown to be on the level of per mille in simulation (Section 4.5). Even if they are underestimated by a factor of two, the correction remains still very small. A conservative systematic uncertainty of 0.5% is assigned.

- p_{T} **extrapolation**. The three analysis methods measure 95-99% of the produced charged particles in the accessible η range (see the left panel of Fig. 4.6). In the tracking method, the minimal p_{T} is below the most probable p_{T}, for the other methods it is even lower. Thus, the extrapolation to $p_{\mathrm{T}} = 0$ is very well constrained: the p_{T} distribution must vanish at zero. The extrapolation to infinite p_{T} is a negligible correction, introduces no uncertainties. Taking into account the minimal p_{T} of the methods, the uncertainties of 0.2%, 0.3%, and 0.5% are assigned for the cluster counting, tracklet, and tracking analysis, respectively.

Source	Pixel Counting [%]	Tracklet [%]	Tracking [%]
Event selection	3.0	3.0	3.0 (1.0)
Detector acceptance	1.0	1.0	1.0
Pixel cluster efficiency	0.5	1.0	0.3
Pixel cluster splitting	1.0	0.4	0.2
Tracklet and cluster selection	3.0	0.5	-
Algorithmic efficiency	-	3.0	2.0
Loopers	2.0	1.0	-
Secondary particles	2.0	1.0	1.0
Misalignment	-	1.0	0.1
Other background hits	1.0	0.2	0.1
Multiple counting, fake rate	-	-	0.5
p_{T} extrapolation	0.2	0.3	0.5
Total, excl. com. uncertainties	4.4	3.7	2.4
Total, incl. com. uncert. of 3.2%	5.4	4.9	4.0 (2.8)

Table 4.6: Summary of systematic uncertainties of the three analysis methods, in percent. In brackets the uncertainties of the average-p_{T} measurement is shown.

5. Measured spectra of unidentified charged hadrons at 0.9, 2.36, and 7 TeV

In this chapter the results on the spectra of unidentified charged hadrons are summarized [53, 54, 55, 56, 57, 58, 59]. First the measured pseudorapidity, $dN_{\rm ch}/d\eta$, and transverse momentum, $dN_{\rm ch}/dp_{\rm T}$, distributions are presented, then they are discussed in the light of earlier experimental findings and various event generator predictions.

5.1 Charged hadron transverse momentum distributions

Tracks with $|\eta| < 2.4$ and $p_{\rm T} > 0.1$ GeV/c were used for the measurement of $dN_{\rm ch}/dp_{\rm T}$. The measured average charged hadron yields per NSD event at 7 TeV are shown in the left panel of Fig. 5.1, as a function of $p_{\rm T}$ for various small $|\eta|$ bins. The yields were fit by the Tsallis function (Eq. 1.31). These fit parameters change by less than 5% with η, thus a fit to the whole region $|\eta| < 2.4$ was also performed. The $p_{\rm T}$ resolution of the CMS tracker (which is smaller than 2%) was found to have a negligible effect on the measured spectral shape and was therefore ignored in the fit function. The role of the fit to the data is twofold: it is used for extrapolations a) to $p_{\rm T} = 0$ (the correction is smaller than 5%); and b) to high-$p_{\rm T}$ ($\sim 1\%$ correction). The fit is not expected to give a precise description of the data in all bins with only two fit parameters. The uncertainties arising from the imperfect fit were taken into account in the systematic uncertainties (see Section 4.6). The yield of charged hadrons, $1/(2\pi p_{\rm T})d^2N/d\eta dp_{\rm T}$, in the region $|\eta| < 2.4$, fit with the Tsallis function is shown in the right panel of Fig. 5.1.

The numerical results are as follows. For the 0.9, 2.36, and 7 TeV data, the inverse slope parameter T and the exponent n of the Tsallis fit were found to be $T = 0.13 \pm 0.01$ GeV and $n = 7.7 \pm 0.2$, $T = 0.14 \pm 0.01$ GeV and $n = 6.7 \pm 0.2$, and $T = 0.145 \pm 0.005$ and $n = 6.6 \pm 0.2$, respectively. The average transverse momentum, calculated from the measured data points adding the low- and high-$p_{\rm T}$ extrapolations from the fit is $p_{\rm T} = 0.46 \pm 0.01$ (stat.) $\pm$ 0.01 (syst.) GeV/c for the

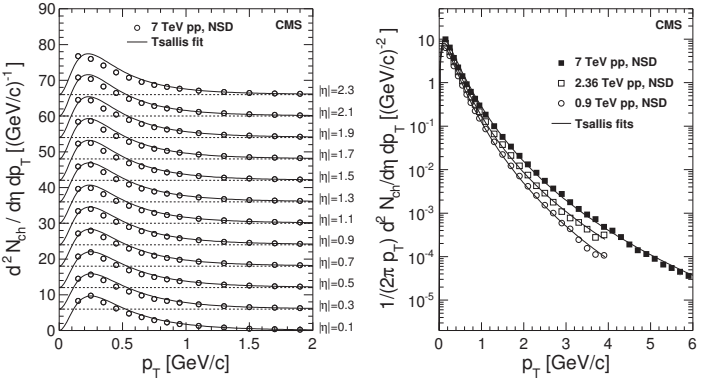


Figure 5.1: *Left*: Measured differential yield of charged hadrons in the range $|\eta| < 2.4$ in 0.2-unit-wide bins of $|\eta|$ for the 7 TeV data. The measured values with systematic uncertainties (symbols) and the Tsallis fit functions are shown. The values with increasing η are successively shifted by six units along the vertical axis. *Right*: Measured yield of charged hadrons for $|\eta| < 2.4$ with systematic uncertainties (symbols) at 0.9, 2.36 and 7 TeV collision energy, fit with the Tsallis function.

0.9 TeV, 0.50 ± 0.01 (stat.) ± 0.01 (syst.) GeV/c for the 2.36 TeV, and 0.545 ± 0.005 (stat.) ± 0.015 (syst.) for the 7 TeV data.

5.2 Charged hadron pseudorapidity densities

The measured pseudorapidity densities of charged hadrons for the three different analysis methods are shown in Fig. 5.2. The $dN_{\rm ch}/d\eta$ results for the three layers in the cluster-counting method and the three layer-pairs in the pixel-tracklet method are consistent within 1.2% and 0.6%, respectively. These results from the various layers and from the different layer pairs were combined to provide one set of data from each analysis method, as shown in the left panel of Fig. 5.2. The error bars include the systematic uncertainties of about $2.4 - 4.4\%$ specific to each method, estimated from the variations of model parameters in the simulation used for corrections and the uncertainties in the data-driven corrections (Section 4.6). The systematic uncertainties common to all the three methods, which amount to

3.7%, are not shown. The results from the three analysis methods are in agreement within uncertainties.

The combined $dN_{\mathrm{ch}}/d\eta$ distributions were calculated as the weighted average of the data from the three reconstruction methods, taking into account their systematic uncertainties, excluding the common ones. The averaged result is shown in the right panel of Fig. 5.2 and compared to measurements at the same accelerator (ALICE, pp [46]) and to previous measurements at the same energy but with different colliding particles (UA5, $p\bar{p}$ [60]). The shaded error bands on the CMS data and the error bars for the data from ALICE indicate systematic uncertainties, while the error bars on the data from UA5 display statistical uncertainties only. The measured $dN_{\mathrm{ch}}/d\eta|_{|\eta|<0.5}$ values for NSD events are 3.48 ± 0.02 (stat.) ± 0.13 (syst.), 4.47 ± 0.04 (stat.) ± 0.16 (syst.), and 5.78 ± 0.01 (stat.) ± 0.23 (syst.) at 0.9, 2.36 and 7 TeV, respectively.

5.3 Discussion

The measured charged hadron multiplicity densities and transverse momentum spectra can be compared to the results of UA1 and UA5 experiments. These experiments were situated at the CERN SPS and recorded data (among other energies) at 0.9 TeV in $p\bar{p}$ collisions. Since at 0.9 TeV the hadronic cross section and the produced hadron densities in the observed $|\eta| < 2.5$ region only differ by a few percent in pp and $p\bar{p}$ collisions, a direct comparison is possible between my results and those obtained at the SPS.

The cross section, $Ed^3\sigma/dp^3$, as a function of p_{T} at 0.9 TeV in $p\bar{p}$ collisions was measured by UA1 [44] (Table 2) together with the charged multiplicity density, $dN_{\mathrm{ch}}/d\eta$ (Table 4). In order to be able to compare their results on cross section as a function of p_{T} to the invariant yields as a function of p_{T} measured at the LHC, a normalization factor is needed to be computed. In other words, both results can be transformed into $(1/2\pi p_{\mathrm{T}})d^2N/d\eta dp_{\mathrm{T}}$ as a function of p_{T} assuming that all particles are pions (see Section 1.3, and in particular the few equations from Eq. 1.23 onwards). However, this procedure can only be followed if the cross section measured by UA1 is inclusive in p_{T}, which is unfortunately not the case: the cross

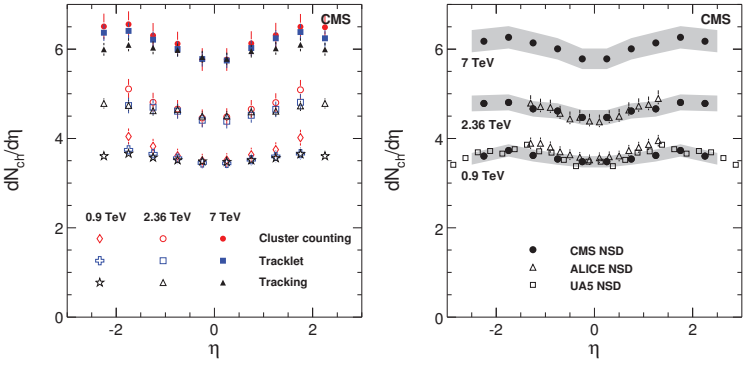


Figure 5.2: *Left*: Reconstructed $dN_{\mathrm{ch}}/d\eta$ distributions obtained from the cluster counting, tracklet and tracking methods at 0.9, 2.36 and 7 TeV. The error bars include systematic uncertainties (as discussed in the text), excluding those common to all the methods. *Right*: Reconstructed $dN_{\mathrm{ch}}/d\eta$ distributions averaged over the cluster counting, tracklet and tracking methods, compared to data from the UA5 [60] (open squares) and from the ALICE [46] (open triangles) experiments at 0.9 TeV. The CMS and UA5 data points are symmetrised in η. The shaded band represents systematic uncertainties of this measurement, which are largely correlated point-to-point. The error bars on the ALICE data points are systematic uncertainties. The error bars on the UA5 data points are statistical only.

section is only measured above 0.2 GeV/c. Thus, an extrapolation is needed to zero transverse momentum. Fortunately, this extrapolation in well constrained due to the published p_{T}-inclusive $dN_{\mathrm{ch}}/d\eta$ results of UA1 (which is the integral of the p_{T} spectrum). The extrapolation is done by assuming a thermal behaviour at low-p_{T}, $\exp(-m_{\mathrm{T}}/T)$ (see Eq. 1.27), and targeting their published $dN_{\mathrm{ch}}/d\eta$ integral. Finally, the comparison of the $(1/2\pi p_{\mathrm{T}})d^2N/d\eta dp_{\mathrm{T}}$ spectra for CMS and UA1 can be seen in Fig. 5.3. The measured values are in very good agreement throughout the p_{T} interval measured by both experiments.

The average transverse momentum $\langle p_{\mathrm{T}} \rangle$ and pseudorapidity density can be compared to earlier experimental results as a function of the collision energy. At low energies the energy dependence of $\langle p_{\mathrm{T}} \rangle$ can be described by a quadratic function of $\ln s$, where s is the collision energy. As shown in Fig. 5.4 the new measurements

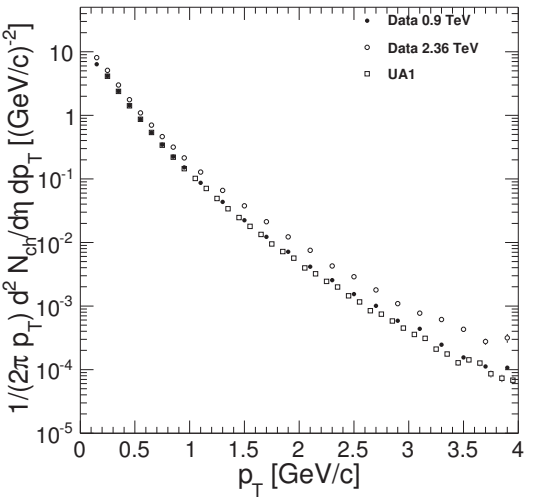


Figure 5.3: Comparison of the measured invariant yields at 0.9 TeV pp (CMS) and $p\bar{p}$ (UA1) collisions. The two set of results are in very good agreement.

follow this trend. In Ref. [16] the authors obtained a power-like collision energy dependence of $\langle p_{\mathrm{T}} \rangle$ which they tuned using experimental data up to 2.36 TeV. The 7 TeV prediction of this model overshoots the measurement. Note that the $\langle p_{\mathrm{T}} \rangle$ values plotted in Fig. 5.4 are computed for various $|\eta|$ intervals. The choice of this interval influences the $\langle p_{\mathrm{T}} \rangle$ results by a few percent ($\langle p_{\mathrm{T}} \rangle$ decreases as the η region of the measurement increases). Our results follow the trend established by the previous experimental results, however, very precise quantitative conclusion is made difficult by the (weak) η dependence of $\langle p_{\mathrm{T}} \rangle$. With the exception of PHOJET, the data is not described by the tested event generators.

The collision energy dependence of the measured $dN_{\mathrm{ch}}/d\eta|_{\eta \approx 0}$ is shown in the left panel of Fig. 5.4, which includes data from the NAL Bubble Chamber [62], the ISR [63], and UA1 [44], UA5 [43], CDF [45], STAR [47], PHOBOS [64] and ALICE [46]. It is worth to mention that the ALICE results are in remarkably good

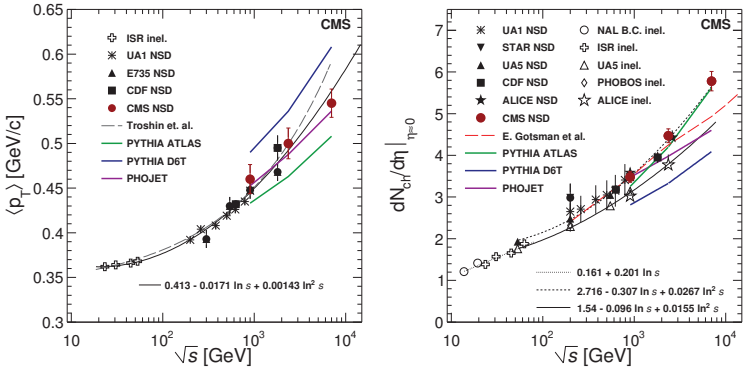


Figure 5.4: *Left*: Energy dependence of the average transverse momentum of charged hadrons. The CMS data points are evaluated for the range $|\eta| < 2.4$. Data of other experiments are taken from Refs. [61, 44, 42, 11]. The curve shows the fit to the data points of the form $\langle p_{\mathrm{T}} \rangle = 0.413 - 0.0171 \ln(s) + 0.00143 \ln^2(s)$ with $\langle p_{\mathrm{T}} \rangle$ in GeV/c and s in GeV^2. Predictions from Ref. [16] and from PYTHIA are also shown as solid lines. The error bars on the CMS data points include systematic uncertainties. *Right*: Charged-hadron pseudorapidity density in the central region as a function of centre-of-mass energy in pp and $p\bar{p}$ collisions including lower energy data from Refs. [62, 63, 43, 44, 45, 46, 47, 64], together with various empirical parameterizations fit to the data corresponding to the inelastic (solid and dotted curves with open symbols) and to the NSD (dashed curve with solid symbols) event selection. Predictions from Ref. [17], from PYTHIA ATLAS and D6T [51] tunes and from PHOJET are also shown as solid lines. The error bars indicate systematic uncertainties, when available.

agreement with ours. The $dN_{\mathrm{ch}}/d\eta|_{\eta\approx 0}$ measurement reported here is consistent with the previously observed trend, it can be well described by a quadratic function of $\ln s$, where s is the collision energy. The energy dependence shows no sign of (even mild) saturation, as is predicted by, for instance, Ref. [17] (red curve). The predictions of the various PYTHIA tunes and the PHOJET generator are overall poor. They can describe neither the magnitude nor the increase of the measured particle densities with the exception of the ATLAS tune, which fits the data relatively well. The predictions on charged hadron density of the various tunes at the three collision energies and the deviations from the measured values are collected in Table 5.1 and

	$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$					
	PYTHIA ATLAS	D6T	DW	Pro-Q20	Perugia-0	PHOJET
0.9 TeV	3.24	2.82	-	2.80	-	3.46
2.36 TeV	4.20	3.31	-	-	-	3.96
7 TeV	5.60	4.07	3.99	4.18	4.34	4.57

Table 5.1: Charged hadron density predictions at midrapidity at various collision energies by various PYTHIA tunes (ALTAS [49], D6T [51], DW [22], Pro-Q20 [23] and Perugia-0 [24]), and the PHOJET [17] event generator.

	$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$(D6T)/$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$(Data)	$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$(PHOJET)/$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$(Data)
0.9 TeV	$0.81 \pm 0.01 \pm 0.03$	$0.99 \pm 0.01 \pm 0.04$
2.36 TeV	$0.74 \pm 0.01 \pm 0.03$	$0.89 \pm 0.01 \pm 0.03$
7 TeV	$0.70 \pm 0.01 \pm 0.03$	$0.79 \pm 0.01 \pm 0.03$

Table 5.2: Charged hadron density ratios at midrapidity. The PYTHIA D6T and PHOJET predictions are compared to the measured values at 0.9, 2.36, and 7 TeV.

Table 5.2, respectively. The numerical values of the increase in charged hadron production is given in Table 5.3.

Looking at both panels of Fig. 5.4 ($\langle p_{\rm T} \rangle$ and $dN_{\rm ch}/d\eta|_{y\approx 0}$ as a function of $\sqrt{s}$) it is clear that none of the event generator tunes can describe both aspects of the data at the same time. On the one hand, PHOJET gives a good description of the evolution of $\langle p_{\rm T} \rangle$, but severely underestimates the hadron production at large collision energies. On the other hand, PYTHIA with the ATLAS tune gives a good description of hadron production but fails to describe the average-$p_{\rm T}$.

The $dN_{\rm ch}/d\eta$ distributions are weakly η-dependent, with a slow increase towards higher η values, and an indication of a decrease at $|\eta| > 2$. The $dN_{\rm ch}/d\eta$

	$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$(2.36 TeV)/$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$(0.9 TeV)	$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$(7 TeV)/$\frac{dN_{\rm ch}}{d\eta}\|_{\eta=0}$(0.9 TeV)
Data	$1.28 \pm 0.01 \pm 0.03$	$1.66 \pm 0.01 \pm 0.02$
D6T	1.19	1.46
PHOJET	1.14	1.32

Table 5.3: The increase of charged hadron density for 0.9 to 2.36 and 7 TeV in data, PYTHIA D6T and PHOJET.

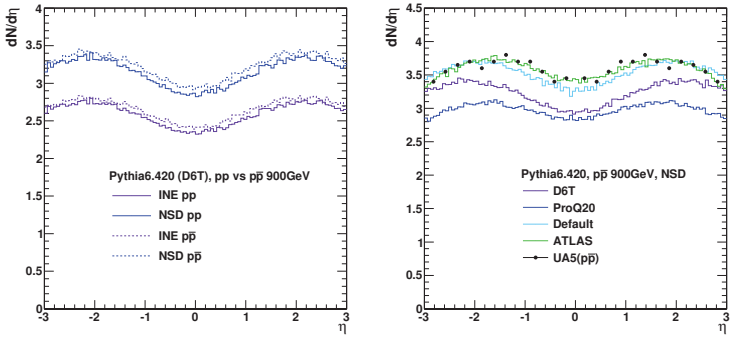


Figure 5.5: *Left*: Comparison of charged hadron multiplicity for NSD (purple) and inelastic (blue) pp and $p\bar{p}$ collisions produced with PYTHIA D6T at 0.9 TeV. Less than 5% difference is seen over the η range of the measurement ($|\eta| < 2.5$). *Right*: Comparison of UA5 with various PYTHIA tunes for 0.9 TeV NSD $p\bar{p}$ collision: most of the PYTHIA tunes underestimate the measurement while the ALTAS tune gives a good match.

distribution at 0.9 TeV is in very good agreement with the measurement of UA5 at the same collision energy with a different collision system, $p\bar{p}$. This is in agreement with our expectation that the inclusive hadron production for pp and $p\bar{p}$ already agree within a few percent at 0.9 TeV. This expectation is reinforced by the results of PYTHIA simulations, see the left panel of Fig. 5.5. The pp and $p\bar{p}$ densities only differ by a few percent.

Since the UA5 data were available since 1986, we could expect that some of the above-mentioned tunes can reproduce it. Indeed, the ATLAS tune of PYTHIA was tuned to reproduce the NSD measurements of UA5, as can be seen in the right panel of Fig. 5.5; the default (or untuned) PYTHIA can also describe this particular observable. Since my measurements are done in pp collisions, the results of the same PYTHIA tunes are also provided for these collisions in the left panel of Fig. 5.6 in comparison with the $p\bar{p}$ results of UA5.

It is worth mentioning that while the NSD $dN_{\mathrm{ch}}/d\eta$ spectrum is well described by the ATLAS tune, neither this one nor any other tune can reproduce the inelastic

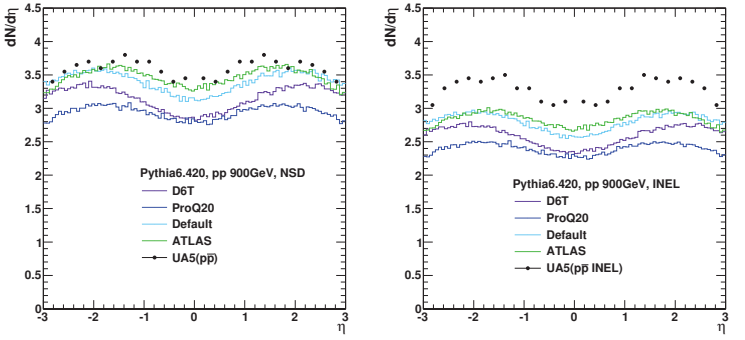


Figure 5.6: Charged hadron multiplicity densities in various PYTHIA tunes for 0.9 TeV NSD (*left*) and inelastic (*right*) pp collision in comparison with $p\bar{p}$ UA5 data. Given that the cross section of charged hadron production in pp and $p\bar{p}$ collisions are similar, the ATLAS tune describes the NSD data well, while all of the PYTHIA tunes (including ATLAS) underestimate the measured yield in inelastic collisions.

$dN_{\mathrm{ch}}/d\eta$ of UA5 ($p\bar{p}$) as shown in the right panel of Fig. 5.6 (the small difference due to the differing pp and $p\bar{p}$ hadron production cross sections is marginal in comparison to the observed difference between the distributions).

This observed difference in the yield of particle production between event generators and data does not mean that QCD in any way is broken in the examined energy regime. Perturbative QCD calculations are only possible at large momentum transfer (where the coupling becomes small), and thus only the high-p_{T} region of the hadron spectra could be described by them. Fig. 5.1 shows that most of the hadron yield comes from the (perturbatively indescribable) low-p_{T} regime. The observed differences between event generators and data (probably) only means that the phenomenological models (which describes the low-p_{T} region) are not well tuned.

This argument is supported by the results of the measurement of hadron spectra up to hundreds of GeV/c by CMS [65]. This was a dedicated measurement done by my colleagues at MIT with the help of jet triggers to collect high-p_{T} particles. From the point of view of my thesis it has two important findings: i) the hadron spectra reconstructed by different vertex finders, different tracking algorithms (both

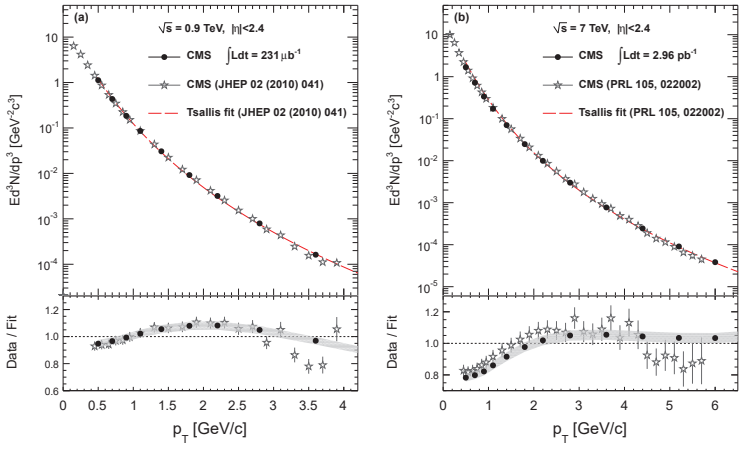


Figure 5.7: Upper panels: Comparison between the invariant charged particle differential yields of this section (stars) and of [65] (solid circles) at 0.9 (*left* panel) and 7 (*right* panel) TeV. Lower panels: The ratio of the two set of results to a Tsallis fit of this section. Error bars on my measurement are the statistical plus systematic uncertainties added in quadrature. The systematic uncertainty band around the new measurement consists of all contributions, except for the event selection uncertainty.

were tuned towards high multiplicity events and the reconstruction of high-p_{T} particles), and different software releases are in good agreement with ours, ii) the large momentum region of the spectrum is in reasonable agreement with the power-law scaling predicted by pQCD (about the scaling see, for instance, Ref. [66]). Figure 5.7 shows the comparison mentioned under "i)" at 0.9 and 7 TeV; the figure is taken from [65]. The power-law scaling of "ii)", to be precise, is the scaling of the inclusive charged particle invariant differential cross section with the variable $x_{\mathrm{T}} \equiv 2p_{\mathrm{T}}/\sqrt{s}$: $Ed^3\sigma/dp^3 = F(x_{\mathrm{T}})/p_{\mathrm{T}}^{n(x_{\mathrm{T}},\sqrt{s})} = F'(x_{\mathrm{T}})/\sqrt{s}^{n(x_{\mathrm{T}},\sqrt{s})}$. By plotting $\sqrt{s}^n Ed^3\sigma/dp^3$ as a function of x_{T} a nice power-law scaling becomes visible for large p_{T} (large x_{T}), as shown in Fig. 5.8 (the figure is taken from [65]).

In the final paragraph of this subsection I would like to add a remark on tuning and the conservation of energy. It is possible to achieve the increase of average-p_{T}

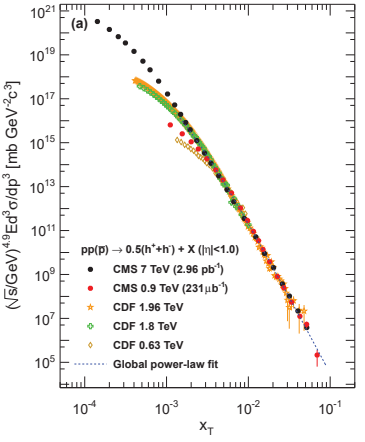


Figure 5.8: Inclusive charged particle invariant differential cross sections, scaled by $\sqrt{s}^{4.9}$, for $|\eta| < 1.0$ (in order to match the acceptance of the measurements by CDF) as a function of the scaling parameter x_{T}. The high-p_{T} (high-x_{T}) power law scaling is clearly visible.

as well as that of the charged hadron density in the central $|\eta| < 2.5$ region at the same time. It does not violate the conservation of energy. The beam momentum is a few thousand GeV/c, which is much larger than the few hundreds of MeV/c of the average-p_{T}. Since my measurements are only done in the central pseudorapidity region, it is possible to "transfer" the beam momentum into a larger amount and a more energetic set of hadrons at midrapidity without putting extra energy into the system "by hand".

5.4 The impact of the results and their afterlife

In this subsection I briefly mention other measurements influenced and inspired by the presented results on multiplicity and transverse momentum distributions.

As I mentioned in Section 4.1.2, the choice of correcting my measurements to NSD collisions was a topic of debate. The ATLAS Collaboration did not take this

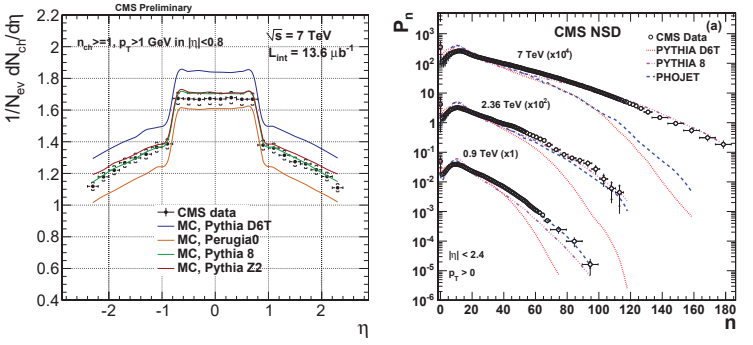


Figure 5.9: *Left*: Charged particle pseudorapidity density distributions in 7 TeV proton-proton collisions. The event selection requires at least one charged particle with $|\eta| < 0.8$ and $p_{\mathrm{T}} > 1$ GeV/c. *Right*: Event-by-event charged hadron multiplicity distributions with $|\eta| < 2.4$ and $p_{\mathrm{T}} > 0$.

approach [67]. They defined a technical event class by requiring at least one track in the $|\eta| < 2.5$ and $p_{\mathrm{T}} > 500$ MeV/c kinematic region. This was favoured by a few event generator developers, most prominently by Peter Skands. On the contrary, the ALICE Collaboration adopted the CMS way of correcting the results to physical collision types [46]. To resolve these issues a cross-experiment working group has been established, called LPCC [68]. It was decided that the experiments continue to report results according to their own decision, but they also have to make the measurements for a set of technically defined collision classes in order to allow direct comparisons between the results from different experiments. As an example to this effort within CMS, the pseudorapidity distribution of charged particles, when there is at least one track with $|\eta| < 0.8$ and $p_{\mathrm{T}} > 1$ GeV/c in the event, is shown in the left panel of Fig. 5.9 [69]. The characteristic edges at $|\eta| \approx 0.8$ are due to the event selection criteria (one track with $|\eta| < 0.8$) mentioned above.

The tuning of generators to describe the observed charged multiplicity started right after the measured results became public. This effort is lead by Richard Field from University of Florida. The recent status of the tuning of PYTHIA is very promising, see for instance, Ref. [70].

The fact, that my measurements show larger charged hadron densities than were previously expected, indicated that the event-by-event multiplicity distribution have a large tail towards large multiplicities. This was already visible in the event-by-event multiplicity distribution of pixel clusters, discussed in Section 4.3.5. A dedicated event-by-event multiplicity measurement [71] was performed using the tracks reconstructed with our minimum bias tracking (instead of the standard tracking). This was another success of our tracking algorithm. The results confirmed the existence of a large multiplicity tail (right panel of Fig. 5.9).

It is an interesting question whether in such a large multiplicity tail, collective heavy-ion collision like phenomena occur or not. In order to enhance the high multiplicity events in the recorded sample of collisions, CMS deployed a special high multiplicity trigger. One of the measurements, which was performed on the high multiplicity pp collision events, is the so-called two-particle correlation analysis [72]. In this analysis the spatial correlation of tracks are computed as a two dimensional distribution of ($\Delta\eta$, $\Delta\phi$), where the difference in pseudorapidity and azimuth angle of tracks are computed for all pairs of tracks. The left panel of Fig. 5.10 shows the two dimensional correlation function for minimum bias (upper row) and large multiplicity (lower row) events in certain p_{T} regions. The correlation function of tracks within $1 < p_{\mathrm{T}} < 3$ GeV/c for high multiplicity events (lower right figure) shows a near side ridge-like correlation, which is not seen on the other figures. This ridge structure has never been seen in proton-proton collisions before, but it was already observed in relativistic heavy-ion collisions [73]. This result was one of the most highlighted ones from CMS and from the whole LHC, and triggered a lot of discussion and articles on the subject.

One advantage of the pixel counting measurement method over the other two, is that it only uses low-level objects, pixel clusters. Thus, it is not affected by the complications, which arise during high-level object reconstruction (for instance, tracking) in a dense environment. The results presented in this section show that the method is reliably applicable to measure the charged hadron densities (since it gives consistent results with those of the more complicated methods) despite of being simple. Hence, an area, where this nature of the method provides good applicability,

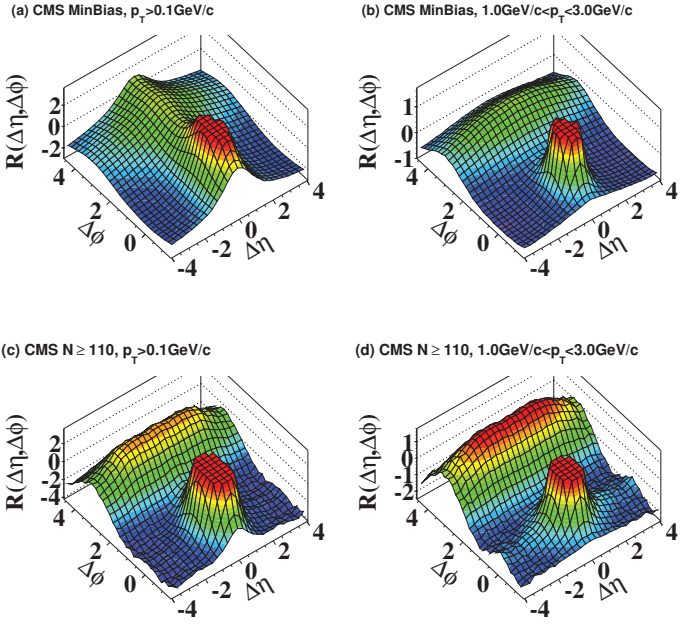


Figure 5.10: Two-particle correlation functions for 7 TeV pp collisions (a) minimum bias events with $p_{\mathrm{T}} > 0.1$ GeV/c, (b) minimum bias events with $1 < p_{\mathrm{T}} < 3$ GeV/c, (c) high multiplicity ($N_{\mathrm{track}} \geq 110$) events with $p_{\mathrm{T}} > 0.1$ GeV/c, and (d) high multiplicity ($N_{\mathrm{track}} \geq 110$) events with $1 < p_{\mathrm{T}} < 3$ GeV/c. The sharp near-side peak from jet correlations is cut off in order to better illustrate the structure outside that region.

is that of heavy-ion collisions. Heavy-ion collisions at a collision energy per nucleon pair of 2.76 TeV were provided by the LHC in the last weeks of November, and in the beginning of December, 2010. I participated in the analysis of the first data by studying the properties of pixel clusters and running the cluster counting analysis. The tracklet analysis was also repeated. I reported our results on the multiplicity density distribution of charged hadrons on the Quark Matter 2011 conference, in

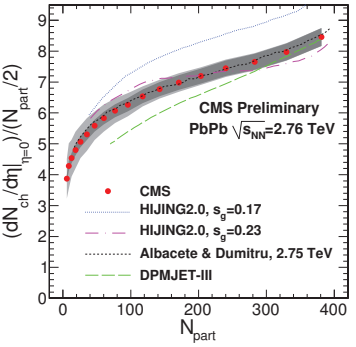


Figure 5.11: Measured $(dN_{\rm ch}/d\eta|_{\eta=0})/(N_{\rm part}/2)$ as a function of the number of participants in 2.76 TeV PbPb collisions [75]. The model predictions are taken from Refs. [76], [77], and [78].

May, 2011 [74, 75]. Figure 5.11 shows the measured charged hadron density normalized by the number of nucleons in the two Pb nuclei which experienced at least one collision, $N_{\rm part}$, as a function of $N_{\rm part}$. The measured distribution is shown together with the predictions of three largely different models [75, 76, 77, 78]. The publication from this measurement is in the last phase of the internal review within CMS, and will be submitted for publication still in July, 2011.

6. Measurement of spectra of strange hadrons

In this section I introduce the methods of measuring the spectra of weakly decaying strange particles (K^0_s, Λ and $\bar{\Lambda}$), so-called *V0*s[26].

These measurements provide a testing ground for the interplay of soft and hard QCD interactions at the LHC, for the universality of fragmentation models between various particles, and for the baryon transport mechanism, which transports the baryon numbers of the incoming proton beams from forward rapidities to central ones. They also provide valuable information for the tuning of Monte Carlo generators.

Two largely independent analysis methods were developed to be performed on data. One by me (*high efficiency* method) [79], the other one by a group from the University of Colorado Boulder (*high purity* method) [80]. I will discuss my method in detail, while the second one will only be briefly mentioned. For the final (published) measurement on data, a combination of these two methods was used.

V0s are long-lived particles ($c\tau > 1$ cm). They can be identified via their decay products originating from a displaced secondary vertex, the decay vertex. The masses, the lifetimes (expressed as $c\tau$), the observed decay channels, and the q values of these channels are listed in Table 6.1.

6.1 High efficiency method

The high efficiency method aims to measure the yields of primary V0s with emphasis on high reconstruction efficiency. The high efficiency comes with the price of a significant amount of V0 candidates due to combinatorial possibilities.

6.1.1 Steps of V0 reconstruction

The reconstruction of V0s is based on the reconstruction of the tracks of their charged daughters.

[26]The name comes from their typical trace in bubble chambers: the neutral mother is unseen, while the traces of the two oppositely charged daughter tracks are bent to different directions by the magnetic field, mimicking the legs of the letter *V*.

Particle	m [GeV/c^2]	$c\tau$	Decay channel (BR)	q [MeV/c]
K^0_s	0.497	2.68	$K^0_s \to \pi^+\pi^-$ (69.2%)	206
$\Lambda/\bar{\Lambda}$	1.115	7.89	$\Lambda \to p\pi^-/\bar{\Lambda} \to \bar{p}\pi^+$ (63.9%)	101

Table 6.1: The masses, the lifetimes (expressed as $c\tau$), the observed decay channels, and the q values of the measured strange particles.

The tracks of daughters are reconstructed with the minimum bias tracking as described in Section 4.5.1. The only difference is that I do not perform the selection of primary tracks in the last step, since the daughters of V0s are not primaries. The final track collection from the three iterative tracking steps is used as inputs for the V0 reconstruction algorithm. Since the track of a charged particle is only reconstructible if it creates a pixel triplet or a pixel pair in the tracker, the decay vertex of a V0 must be inside the second pixel barrel layer (in a cylinder around the beamline with a radius of ~ 7 cm), in order to be reconstructed. As a consequence, high-p_{T} V0s will have small (but well determined) reconstruction efficiency.

The V0 finder In the volume of the pixel detector the magnetic field is practically constant, charged particles propagate on helices. The search for V0 candidates reduces to the determination of the closest point of the two trajectories of daughters. The applied closest-point-search algorithm calculates the following quantities:

- Transverse distance, Δr, and azimuthal angles, ϕ_1 and ϕ_2, of the closest points of the transverse projections of the two trajectories. The closest points then are mapped back to three dimensional helices[27]. Note that in case of intersecting transverse projections two pairs of points are returned.
- The longitudinal distance, Δz, of the closest points.
- The positions of the closest points, V_1 and V_2. In case of two pairs of closest points the one with smallest Δz is chosen.
- The momenta $\vec{p}_1$ and $\vec{p}_2$ of the two tracks at V_1 and V_2.

[27]The tracking algorithm is such that it either propagates the trajectories from inwards to outwards, or vice versa. Thus, tracks which are completely bend in such a way that they remain in the Tracker Detector are only tracked until the point of "turn over".

The position of the decay vertex is approximated by the midpoint of the line segment V_1V_2. Let $\vec{r}$ be the position vector of such a midpoint calculated with respect to the beamspot. The momentum vector, $\vec{p}$, and the impact parameter, b, of the mother particle can now be computed as follows:

$$\vec{p} = \vec{p}_1 + \vec{p}_2,\ b = \left| \vec{r} - \frac{\vec{p}(\vec{p}\vec{r})}{|\vec{p}|^2} \right|. \tag{6.1}$$

A neutral mother particle is only formed if the two daughter tracks have opposite electric charge. If there are n reconstructed tracks in an event, the number of such V0 daughter candidate pairs can be estimated with $(n/2)^2$. Hence, for high multiplicity events the number of combinations becomes enormous. Therefore, it is important to properly filter tracks and track-pairs in order to speed up the computation and to reduce the fraction of background candidates.

In order to suppress the combinatorial background, various cuts are applied during the V0 reconstruction. The quantities, which are used to suppress the background are the following: the minimal impact parameter of the daughter tracks (b_{pos}, b_{neg}), the maximal impact parameter of the mother particle (b), the maximal three dimensional distance of the trajectories of daughters at the points of closest approach (d), and the minimal and maximal transverse distance of the decay vertex from the beamspot (r_{min}, r_{max}). The signal over background ratio in the sample of V0 candidates increases as b_{pos} and b_{neg} gets higher (primary particles have smaller impact parameter than secondaries), or b and d gets smaller (the mother is a primary particle, the daughters are produced in a point-like decay vertex). Note that in this reconstruction method I favour high efficiency (keeping a large fraction of the signal) over high purity (increasing the signal/background ratio even if part of the signal is sacrificed). Hence, I will keep the above selection requirements as loose as possible in the following sense. The selection parameters cannot be loosened above a certain practical limit due to the exponential increase in the background fraction, which makes the extraction of the signal more and more challenging.

The values of the parameters of the V0 finder was optimized using simulated events, resulting in the numbers given in Table 6.2.

Notation	Meaning	Cut [cm]
b_{pos}	Minimal impact parameter of the positive daughter	0.0
b_{neg}	Minimal impact parameter of the negative daughter	0.0
b	Maximal impact of the mother	0.5
d	Maximal 3D distance of daughter trajectories	0.2
r_{min}	Minimal transverse distance of the decay vertex	0.1
r_{max}	Maximal transverse distance of the decay vertex	7.0

Table 6.2: The V0 selection parameters and their applied values used in the high efficiency analysis.

The $\Lambda/\bar{\Lambda}$ particles are produced in less abundance than K^0_s. Thus, the extraction of their yields is more challenging due to the smaller signal over background ratio. In order to suppress the combinatorial background, a further selection is applied. This selection exhibits the fact that the positive (negative) daughter in the observed decay channel of Λ ($\bar{\Lambda}$) is a proton (antiproton). The fraction of V0 candidates, where the daughter is indeed a p can be enhanced by using the deposited energy per unit path length, dE/dx, in the Tracker Detector [81, 41]. The used dE/dx parameterization is the following:

$$dE/dx = \log(a/\beta^2 + b\log(\beta\gamma) + c), \tag{6.2}$$

where $a = 3.25$, $b = 0.288$, $c = -0.852$, $\beta\gamma = p/m$, and β can be expressed with $\beta\gamma$, $\beta = \beta\gamma/\sqrt{(\beta\gamma)^2 + 1}$. The variance of the dE/dx measurement can be parametrised as a function of the number of the hits on tracks, $\sigma(n) \approx 0.3 \cdot n^{-0.65}$. In the analysis a gentle dE/dx compatibility of 6σ with the proton curve was required for $p/\bar{p}$ candidates.

Extraction of particle yields In order to extract the yields of V0s the distributions of invariant mass,

$$M^2 = (E_1 + E_2)^2 - |\vec{p}_1 + \vec{p}_2|^2 = m_1^2 + m_2^2 + 2(E_1E_2 - (\vec{p}_1\vec{p}_2)), \tag{6.3}$$

are calculated for K^0_s and $\Lambda/\bar{\Lambda}$ with the appropriate $\pi\pi$ or $p\pi$ daughter mass hypothesis. The invariant mass distributions are calculated in two dimensional rapidity-transverse momentum ($y - p_{\mathrm{T}}$) bins. First, I will show the integrated mass

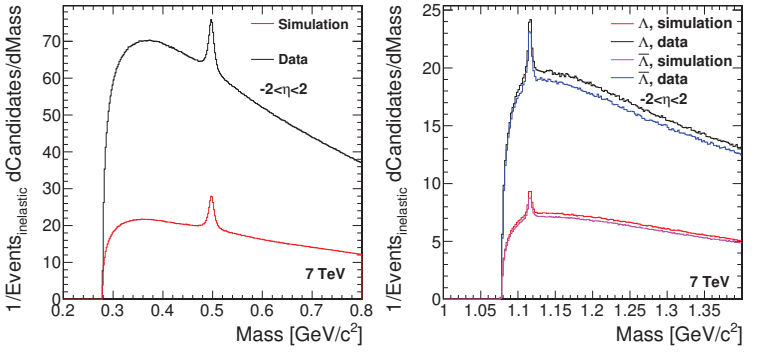


Figure 6.1: Invariant mass distributions for K^0_s (*left*) and $\Lambda/\bar{\Lambda}$ (*right*) at 7 TeV in PYTHIA D6T simulation and in data. The larger amount of background in data than in the D6T simulation is due to the larger charged hadron multiplicity, as discussed in Section 5

.

distributions, and then I will give four examples for invariant mass distributions in $y - p_{\mathrm{T}}$ bins.

The p_{T}-inclusive normalized invariant mass distribution in the $-2 < \eta < 2$ region for the 7 TeV data and PYTHIA D6T simulation are shown in Fig. 6.1. The figure shows three noticeable features. At the first glance the most striking feature is the large difference in the level of background. This feature is due to the larger number of charged particles in data than in PYTHIA D6T simulation (which was one of the results presented in Section 5). In the high efficiency analysis (due to the loose V0 candidate selection cuts) the level of background is sensitive to the initial track multiplicity. The second feature is the different shape of the background in data and in simulation. This has two separate origins: a) the one-particle momentum distribution is different; b) the two-particle correlations are different in data and simulation. The third feature is that the signal, the mass peak itself is larger in data than in simulation. The precise determination of the increase in the yields of strange particles observed in data is one of the objectives of this measurement.

To extract the signal of strange hadrons (the area of the mass peak) from the

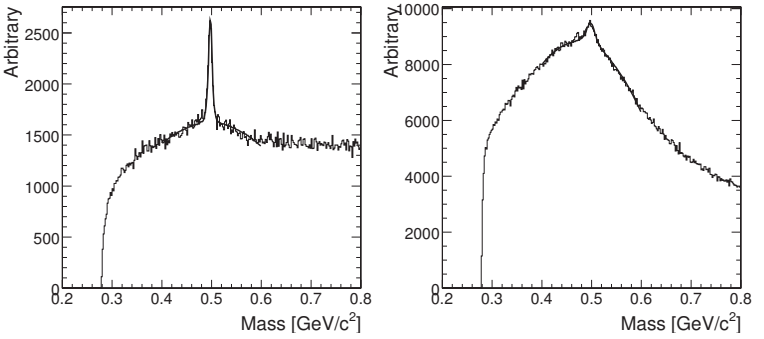


Figure 6.2: Invariant mass distributions with K^0_s hypothesis in the $y \approx -1.1$ and $p_{\mathrm{T}} \approx 1.3$ GeV/c (*left*) and in the $y \approx -0.5$ and $p_{\mathrm{T}} \approx 0.3$ GeV/c (*right*) bins in the 7 TeV data. The fits around the mass peaks are also shown.

invariant mass distribution, I perform a fit as follows: a) I take the shape of the mass peak from the simulation; b) I fit the background with a parabolic function. This strategy was the result of various tests: I tried various fit functions for the mass peak and also for the background. The mass peak was not well reproduced in all $y - p_{\mathrm{T}}$ bins simultaneously by any of the common fit functions (Gauss, Cauchy). This lead to the usage of the shape of the mass peak from the simulation, which also matches the peak in data data nicely. The function used to describe the background below the mass peak was chosen to be the simplest one which still fits the background in all bins. The invariant mass distributions in four $y-p_{\mathrm{T}}$ bins together with the mass fits are shown in Fig. 6.2 and Fig. 6.3. For these figures the rapidity and transverse momentum bins were chosen to show examples for very different mass distributions. All of them can be adequately described by the fitting strategy described above, which uses only four fit parameters: one for the magnitude of the mass peak and three for the shape of the background.

6.1.2 Corrections

In Section 4.5.2 in order to examine the performance of the charged particle tracking, I introduced various notions, such as acceptance, efficiency, etc. In the

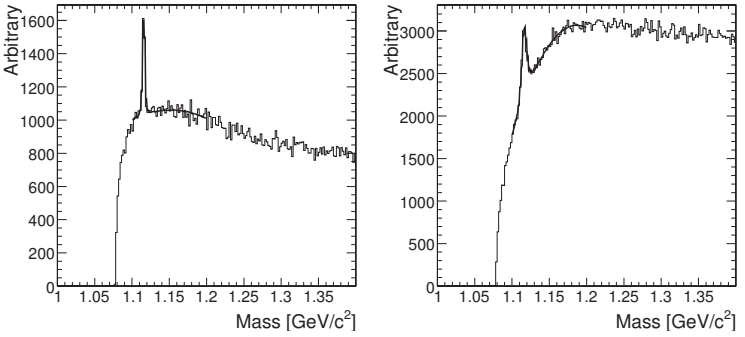


Figure 6.3: Invariant mass distributions with $\Lambda/\bar{\Lambda}$ hypothesis in the $y \approx -1.1$ and $p_{\mathrm{T}} \approx 3.0$ GeV/c (*left*) and in the $y \approx -1.1$ and $p_{\mathrm{T}} \approx 0.45$ GeV/c (*right*) bins in the 7 TeV data. The fits around the mass peak are also shown.

following I will partially extend those definitions to V0s, and I will introduce some new ones, too.

- A V0 decaying in one of the measured decay channels (reported in Table 6.1) is *accepted* if its daughters (which are charged particles) are accepted.
- The *acceptance* is the fraction of the accepted V0s.
- A V0 is *reconstructible* if its daughters are reconstructed by the tracking algorithm as tracks.
- The *reconstructibility* is the fraction of those accepted V0s which are reconstructible.
- The *efficiency* is the fraction of those reconstructible V0s found by the V0 finder algorithm (so which are actually reconstructed).

The acceptance as a function of y and p_{T} can be seen in Fig. 6.4. Note that according to the definition of accepted V0s, the branching ratio is included in the acceptance. Thus, the acceptance is limited by the branching ratios (69.2% for K^0_s and 63.9% for Λ). Apart from this limit, the acceptance can be thought as a

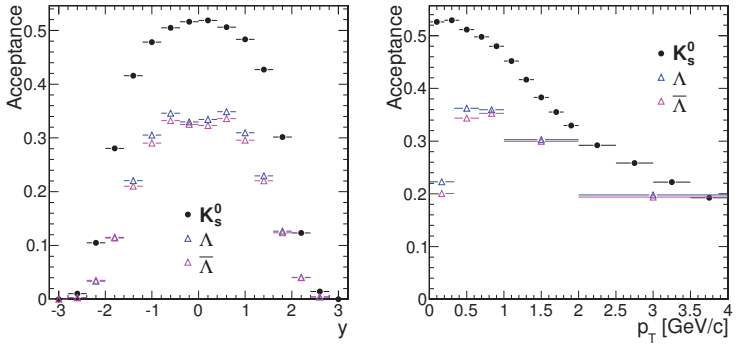


Figure 6.4: Acceptance as a function of y (*left*), and as a function of p_{T} ($|y| < 1.6$) (*right*) for K^0_s and $\Lambda/\bar{\Lambda}$.

coincidence single charged particle acceptance for the two daughters of a V0. It changes slightly as a function of y for $|y| < 0.8$, then it decreases and reaches zero around $|y| \approx 2.5$. At low-p_{T} the acceptance rises (just for single charged particles) then at high-p_{T} it decreases. This is due to the fact that the lifetime observed in the laboratory increases with γ ($t' = \gamma t$, where t' is the observed and t is the rest frame lifetime). Thus, as the particle has larger and larger p_{T} it will be less and less likely to decay in the allowed 7 cm vicinity of the beamline (Table 6.2).

Figure 6.5 shows the reconstructibility as a function of y and p_{T}. The reconstructibility measures the possibility of coincidental reconstruction of daughter tracks. It is larger for K^0_s at mid-rapidity due to larger overall single track reconstruction efficiency for pions as opposed to protons. For larger rapidities the three sets of points get closer and the reconstructibility increases due to the corresponding larger momentum and thus smaller multiple scattering. At low-p_{T} K^0_s and $\Lambda/\bar{\Lambda}$ shows different behaviour: the reconstructibility for K^0_s starts to decrease then it increases, while for $\Lambda/\bar{\Lambda}$ it increases starting from the very first p_{T} bin. These different behaviours are due to the different q values of the observed decay channels, $K^0_s \to \pi^+\pi^-$ and $\Lambda \to p\pi^-/\bar{\Lambda} \to \bar{p}\pi^+$ (Table 6.1). The large q for K^0_s means that one of the daughter pions can be boosted to very small p_{T}, where it falls below the

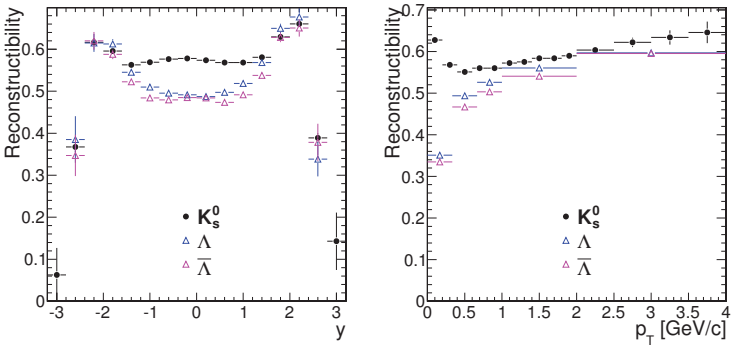


Figure 6.5: Reconstructibility as a function of y (*left*), and as function of p_{T} ($|y| < 1.6$) (*right*) for K^0_s and $\Lambda/\bar{\Lambda}$.

p_{T} limit of track reconstruction. At high-p_{T} the boost provided by q is to small to be able to stop one of the pions, so the reconstructibility increases as p_{T} increases in case of the single charged particle reconstruction efficiency (Section 4.5.2).

In Fig. 6.6 the efficiency is shown as a function of y and p_{T}. According to the definition of the efficiency, it measures the performance of the V0 finder itself. It is the fraction of the reconstructed V0s out of the potentially reconstructible ones (those which have their daughters reconstructed). The efficiency as a function of y is around 88-90% thanks to the open V0 finder cuts. It is flat and slightly higher for K^0_s. The efficiency increases as a function of p_{T}; it becomes 1 for K^0_s and ~ 0.9 for Λ above ~ 1 GeV/c. The efficiency differs by 10% at high-p_{T} for K^0_s and Λ, but the p_{T}-integrated efficiencies are very close as a function of y. The reason for this is that the average-p_{T} of Λ is much higher than that of K^0_s in the simulation. This is confirmed by the average-p_{T} measurement performed in data, as will be shown in Section 7.

There is still a correction needed for the final fully corrected NSD spectra of strange hadrons. That is the event selection efficiency. The correction for imperfect NSD event selections is done in the same way as in the tracking method described in Section 4.5.2.

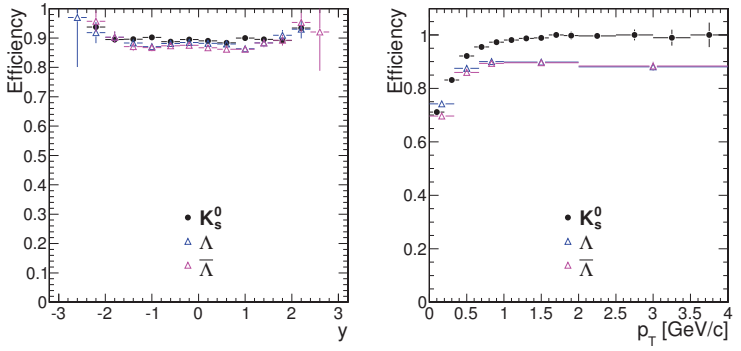


Figure 6.6: Efficiency as a function of y (*left*), and as function of p_{T} ($|y| < 1.6$) (*right*) for K^0_s and $\Lambda/\bar{\Lambda}$.

6.1.3 Podolanski-Armenteros variables

The Armenteros variables are commonly used for parameterizing the V0 candidates.

The momenta of the daughters, p_1 and p_2, can be decomposed to components parallel and perpendicular to the momentum of the mother, $\vec{p} = \vec{p_1} + \vec{p_2}$. The longitudinal components can be obtained by

$$p_{1L} = \vec{p}\vec{p_1}/|\vec{p}| \text{ and } p_{2L} = \vec{p}\vec{p_2}/|\vec{p}|. \tag{6.4}$$

Then, the Armenteros variables q_{T} and α are defined by

$$q_{\mathrm{T}} = |\vec{p_1} \times \vec{p_2}|/|\vec{p}| \tag{6.5}$$

and

$$\alpha = (p_{1L} - p_{2L}) / (p_{1L} + p_{2L}) = \left(\vec{p_1}^2 - \vec{p_2}^2\right) / \vec{p}^{\,2}. \tag{6.6}$$

Using these variables the K^0_ss and Λs are well distinguishable, as shown in Fig. 6.7 for simulated (figure on the left) and selected real collision (figure on the right) events. The band of K^0_s is the large arc ranging between $|\alpha| \approx 0.8$. The two smaller arcs are Λs (right arc) and $\bar{\Lambda}$s (left arc). The larger background in

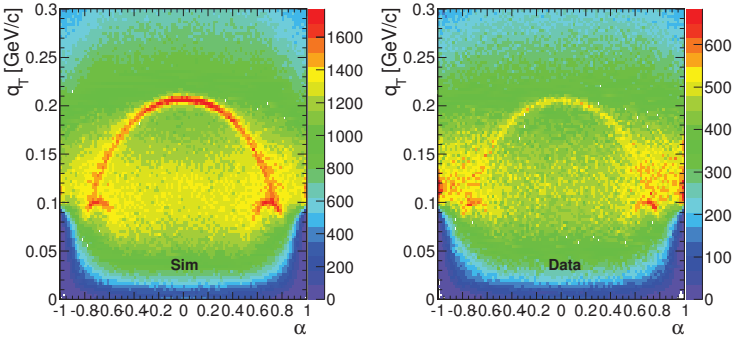


Figure 6.7: Armenteros plot at 900 GeV in simulated (*left*) and selected real collision events (*right*). Bands of K^0_ss, Λs, and $\bar{\Lambda}$s are well visible.

data compared to simulation (already observed earlier) is well visible in the figures. The Armenteros plots suggests that the background can be suppressed by cutting on the arcs in the Armenteros plane. This is true; however it is not practical to use. Cutting on the Armenteros variables means implying a "filter" on the invariant mass distribution of the V0 candidates. This filter leads to the distortion of the background shape in the invariant mass distribution. Such a filter makes the description of the background more complicated. Since in the analysis the background is fit, it is preferable to have featureless, easily fittable background over a more complex one, even if it comes at the price of increasing the overall background fraction. Thus, in this analysis no selection was performed on the V0 candidates based on the Armenteros variables.

6.1.4 Validation of the reconstruction algorithm on simulated data

In order to demonstrate that the reconstruction algorithm is capable of reconstructing the "real" spectra of V0s, simulated events were subjected to the full analysis chain. In such events the simulated and the reconstructed distributions can be compared, thus their agreement can be inspected. Simulated events are treated

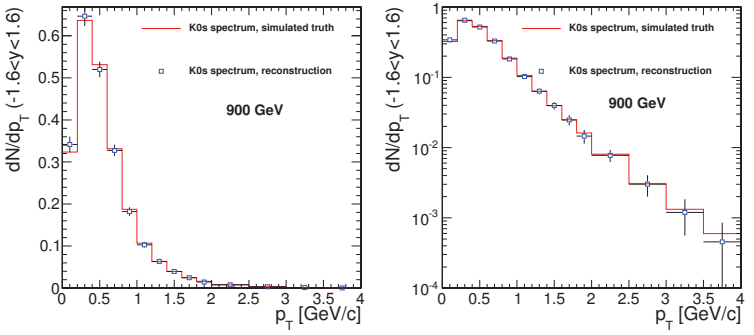


Figure 6.8: Simulated and reconstructed yields of K^0_s in the $|y| < 1.6$ region on a linear (*left*) and a logarithmic scale (*right*).

as if they would be real collision events: the same trigger conditions are required[28] and the same reconstruction algorithms are used as on collision data.

The simulated and reconstructed p_{T} distributions of K^0_s are shown in Fig. 6.8. The good agreement between the simulation and reconstructed distributions give us confidence that the reconstruction method works properly. For Λ the same level of agreement is observed between data and simulation.

The other distribution I aimed to measure is that of the rapidity. The dN/dy distributions are computed by integrating the $d^2N/dydp_{\mathrm{T}}$ distributions (computed in small bins of rapidity) in y. The reconstructed $d^2N/dydp_{\mathrm{T}}$ distributions together with the obtained and simulated dN/dy distributions are shown in the left panel of Fig. 6.9. The pink, black, red and blue symbols show the reconstructed dN/dp_{T} distributions in various y bins; the green histogram is the simulated input dN/dy distribution while the green symbols show the reconstructed dN/dy distribution. As in the case of the transverse momentum distributions shown in the previous figures (Fig. 6.8), the agreement between the simulated and the reconstructed distributions is very good.

The measurement can also be performed using the pseudorapidity variable

[28] Except the BPTX coincidence (which is always true in this case: the event generators only simulate collision events).

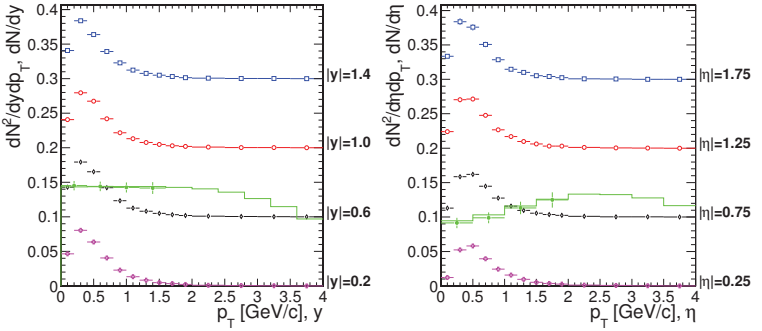


Figure 6.9: *Left*: Reconstructed differential yields, $d^2N/d\eta dp_{\rm T}$, of K^0_s in various y intervals. The simulated and reconstructed dN/dy distributions are also shown by the green histogram and green symbols, respectively. *Right:* Differential yields and hadron density distributions using pseudorapidity instead of rapidity.

instead of rapidity. This is demonstrated in the right panel of Fig. 6.9, which shows the reconstructed $d^2N/d\eta dp_{\rm T}$ and the reconstructed and simulation $dN/d\eta$ distributions. Again, a good agreement is observed between the reconstructed and simulated distributions. For the actual measurement I chose to use the rapidity variable.

6.2 High purity method

In this analysis method the triggering and reconstruction strategy differs significantly from the ones used for the high efficiency analysis. Thus, it provides largely independent results. In the next paragraph I will briefly introduce the high purity method focusing on its peculiarities.

The large-multiplicity beam background filter was completely different than in the previous analysis (described in Section 4.1.2). The filter used here requires a few tracks to be of high "quality", meaning to have small χ^2, and also a reconstructed vertex. The alternative filter, used by the high efficiency analysis, the cluster-vertex compatibility filter, only uses the position and the shapes of pixel clusters.

The tracking algorithm applied in this analysis was the standard CMS track-

ing, which was (despite of its intense development partially motivated by our minimum bias tracking) still inferior in reconstructing low-p_{T} particles with respect to the minimum bias tracking, described in Section 4.5. It still used a straight line "third hit prediction" instead of the exact geometrical solution.

However, it also had a advantage at the same time: it utilized strip seeds as well as pixel seeds to start the track reconstruction. These strip seeded sequences were added to the iterative tracking scheme as further reconstruction steps. The strip seeding steps allow to reconstruct the daughters of V0s even if the decay happened relatively far from the beam line, but they also increase the fake rate significantly (to be at the percent level).

In the high purity analysis the applied vertexing algorithm is a "divisive" finder. It scans through the z-axis starting from one side of the interaction point and divide the region into subregions according to the compatibility of track positions. This is the standard vertexing procedure, but it is shown to be inferior to that of the agglomerative finder (Section 4.2) [82, 48].

The V0 finder itself was also different. It worked as follows. The V0 candidates were identified by selecting pairs of oppositely charged tracks, fit to a displaced secondary vertex. The tracks were reconstructed from the inner tracker, requiring a χ^2/ndf for the track fit less than 5. Secondary tracks were selected requiring a distance of closest approach to the PV in 3 dimensions greater than 3 standard deviations (σ). The secondary vertex (SV) was also required to be separated by more than 5σ from the primary vertex, where σ included the uncertainty from both the PV and SV. The SV was also required to be located not more than 4σ further from the PV than the innermost hit of each of the two daughter tracks, to ensure the tracks were consistent with having originated at the secondary vertex.

In case of the reconstruction of K^0_s, both charged particles were assumed to be pions and the invariant mass of each pair was reconstructed. In order to suppress the contamination at high-p_{T} due to $\Lambda \to p\pi^-$ decays, the corresponding invariant mass of the track pair using a $p\pi^-$ hypothesis was required to be 2.5σ greater than the Λ mass value. Only prompt K^0_s were selected by requiring its momentum vector to point back to the primary vertex within 3σ. The number of K^0_s candidates was

extracted by fitting the data with the sum of a double Gaussian for the signal and a second-order polynomial function for the background.

During the reconstruction of Λs, the contamination due to K_s^0 was suppressed by rejecting candidates where the $\pi^+\pi^-$ pair hypothesis resulted larger mass than the K_s^0 mass allowing a deviation of 2.5σ. The number of Λ candidates was extracted by fitting the data with the sum of a double Gaussian for the signal, and a function of the form $Aq^B e^{Cq+Dq^2}$, where $q = m_\Lambda - (m_p + m_\pi)$, for the combinatorial background.

This reconstruction method has a smaller background fraction, but also a smaller efficiency compared to the high efficiency method.

6.3 Systematic uncertainties

The systematic sources affecting the two analyses are examined in this section. The various sources of systematic uncertainty are discussed separately, then they are summarized in Table 6.3. The quoted systematic uncertainties refer to the measurement at 7 TeV.

- **Event selection**. The high efficiency analysis uses the same event selection as the already discussed charged hadron multiplicity measurement. The high purity analysis uses almost the same triggering strategy; the only difference is the beam background removal tool and the vertexing algorithm (used for the "at least one reconstructed vertex" criterion). Nevertheless, the systematic uncertainties are the same: 3.5%. Note that in the case of the charged hadron spectra I quoted an uncertainty of 3.0% (Section 4.6). There is no inconsistency between these numbers: the numerical values of the uncertainties mentioned in Section 4.6 applied to the 2.36 TeV data, while here the uncertainties apply to the 7 TeV data.

- **Track reconstruction**. The high efficiency analysis uses the same tracking algorithm as the *tracking method* in case of the unidentified spectra measurements (see Section 4.5). Thus, the tracking related systematic uncertainties discussed in Section 4.6 apply here, too (2.5%). The high purity analysis uses a different kind of track reconstruction but the systematic errors are estimated

to be similar.

- **Selection of V0s**. In both analyses the systematic uncertainty associated to V0 reconstruction comes primarily from the selection of V0 candidates. In case of the high purity analysis, the matching criteria between simulated and reconstructed V0s (this association is used during the computation of the correction factor) are looser than for the high efficiency analysis. This introduces an additional systematic uncertainty of 1%. The uncertainty is estimated to be 2.5% (5.0%) for K_s^0 and 4% (5.5%) for Λ in the high efficiency (high purity) analysis by repeated analyses using various selection criteria (an example is shown in Section 6.3.1).

- **Mass fits**. Both analyses rely on invariant mass fits to extract the numbers of V0s per bin. These numbers are then subjects of corrections. The actual mass range of the fit has a small effect on the extracted yield. The effect is 2% for the particle yield of K_s^0 and 4% for Λ in the high efficiency method, negligible in the high purity method. The average-p_{T} is only affected by 1%.

- p_{T} **extrapolation**. In case of K_s^0 the p_{T} spectrum is measured below the maximum of the p_{T} distribution with good precision in both analyses. Thus the extrapolation to 0 is well constrained. The extrapolation uncertainty is estimated to be 0.5%. In case of Λ (due to the significant background) the high efficiency analysis has large uncertainties in the very first bin (the uncertainties are smaller in the high purity analysis). The effect of this problematic bin is suppressed due to the constraint of the other measured points resulting in an uncertainty of 3% (1.8%) for the high efficiency analysis (high purity).

6.3.1 Increase purity in the high efficiency analysis

In order to test the sensitivity of the results to the cuts of the V0 finder in the high efficiency analysis, the analysis was repeated with stricter cuts. This way the background in the invariant mass distributions can be suppressed. However, this comes at the price of decreased reconstruction efficiency. The values of the stricter and the nominal cuts applied in the analysis are listed in Table 6.4.

Source	High efficiency [%] K^0_s	High efficiency [%] Λ	High purity [%] K^0_s	High purity [%] Λ
Event selection	3.5	3.5	3.5	3.5
Track reconstruction	2.5	2.5	2.5	2.5
Selection of V0s	2.5	4.0	5.0	5.5
Mass fit ranges	2.0 (1.0)	4.0 (2.0)	-	-
p_{T} extrapolation	0.5	3.0	0.5	1.8
Total uncert.	5.4 (5.1)	7.7 (6.8)	6.6	7.2

Table 6.3: Summary of the systematic uncertainties affecting the two V0 analysis methods, in percent. In brackets the errors of the average-p_{T} measurement is shown.

Parameters	Default cut [cm]	Stricter cut [cm]
minImpactPositiveDaughter	0.0	0.2
minImpactNegativeDaughter	0.0	0.2
maxDca	0.2	0.2
minCrossingRadius	0.1	0.2
maxCrossingRadius	7.0	7.0
maxImpactMother	0.5	0.2

Table 6.4: The default and the stricter cuts of the V0 finder used in the high efficiency analysis.

These new cuts result in a highly increased signal over background ratio, as shown in Fig. 6.10 for K^0_s (left) and Λ (right). The remaining background peak at the lower edge of the invariant mass distributions comes from photon conversions.

The results of the repeated analysis with the aforementioned stricter cuts are in good agreement with the results of the nominal analysis, as shown in Fig. 6.11 and Fig. 6.12 for K^0_s and Λ, respectively.

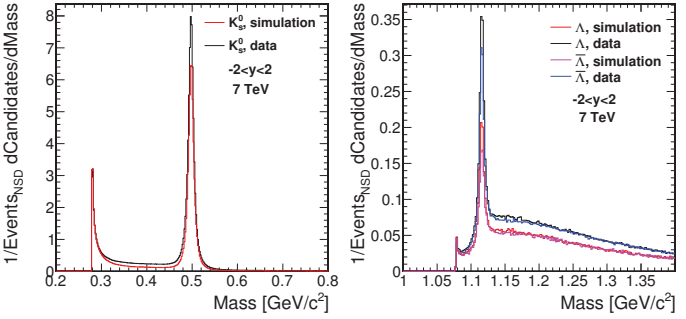


Figure 6.10: Invariant mass distributions with K^0_s (*left*) and Λ (*right*) mass hypotheses at 7 TeV. The distributions are computed in the $|y| < 2$ interval using the stricter V0 finder cuts.

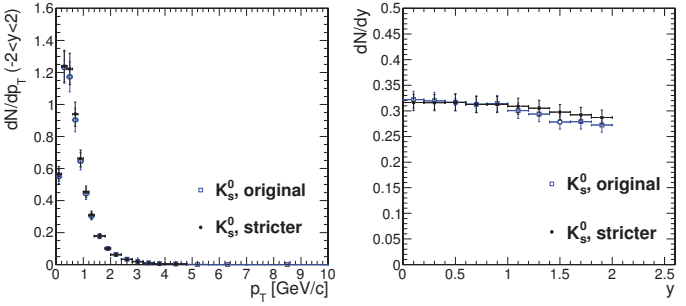


Figure 6.11: Measured dN/dp_{T} (*left*) and dN/dy (*right*) distributions of K^0_s at 7 TeV using the default (open symbols) and the stricter (filled symbols) V0 selection cuts.

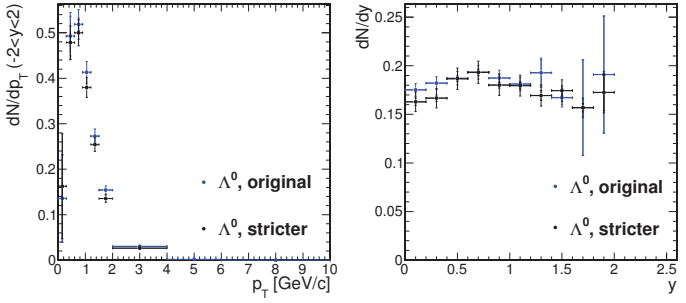


Figure 6.12: Measured dN/dp_{T} (*left*) and dN/dy (*right*) distributions of Λ at 7 TeV using the default (open symbols) and the stricter (filled symbols) V0 selection cuts.

7. Measured spectra of strange hadrons at 0.9 and 7 TeV

The obtained results on the spectra of strange hadrons using the high efficiency method is summarized here. In the figures shown in this section, the inner error bars represent "statistical-like" uncertainties (fit uncertainties propagated from the fits of the invariant mass distributions) while the outer error bars show the quadratic sum of statistical and systematic uncertainties. The only exception is that of the Discussion section, Section 7.3. In that section the results which are compared to the model predictions and earlier experimental results are from a combined analysis of the high efficiency and high purity methods. The aim of the combined analysis was to use the superior features of the two methods omitting the inferior ones [83, 59, 84, 85]. This will be briefly discussed in the beginning of the Discussion section.

7.1 Transverse momentum distributions

In the following the measured values are given in the form of *value* ± *fit uncertainties* ± *systematic uncertainties*. The fit uncertainties represent the statistical uncertainties propagated from the uncertainty of the fits of the invariant mass distributions.

I used tracks with $|y| < 2$ and $p_{\mathrm{T}} > 100$ GeV/c for the measurement of transverse momentum spectra. Outside of this kinematic region the acceptance is small, and the fits of the invariant mass distributions have large uncertainties. The obtained transverse momentum distributions at 0.9 and 7 TeV are shown in Fig. 7.1 for K^0_s) and Λ. The obtained Tsallis fit parameters are the following: a) K^0_s: $n = 7.77 \pm 0.5 \pm 0.04$, $T = 0.177 \pm 0.005 \pm 0.01$ GeV at 0.9 and $n = 6.96 \pm 0.07 \pm 0.05$, $T = 0.216 \pm 0.002 \pm 0.02$ GeV at 7 TeV; b) Λ: $n = 10.02 \pm 1.4 \pm 0.05$, $T = 0.229 \pm 0.01 \pm 0.01$ GeV at 0.9 and $n = 9.01 \pm 0.43 \pm 0.4$, $T = 0.289 \pm 0.005 \pm 0.01$ GeV at 7 TeV. The measured average-p_{T} values computed from the Tsallis fit are the following: $\langle p_{\mathrm{T}} \rangle = 0.63 \pm 0.01 \pm 0.03$ and $0.78 \pm 0.01 \pm 0.02$ GeV for K^0_s and $0.87 \pm 0.03 \pm 0.05$ and $1.04 \pm 0.03 \pm 0.05$ GeV for Λ.

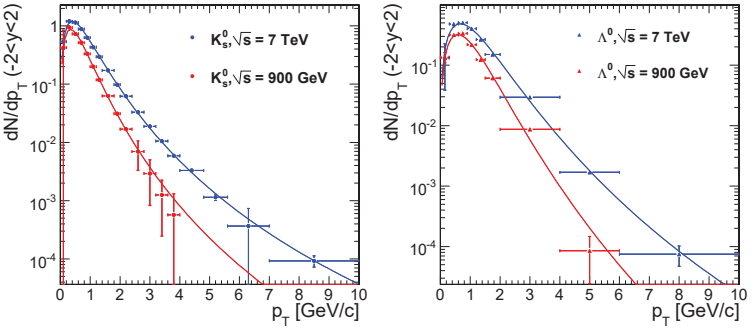


Figure 7.1: Measured dN/dp_{T} distribution of K^0_s (*left*) and Λ (*right*) at 0.9 and 7 TeV in the $|y| < 2$ range. Results are corrected to NSD collisions.

7.2 Pseudorapidity density

The dN/dy distribution is computed by integrating the $dN^2/dydp_{\mathrm{T}}$ distributions in p_{T}. Since the measured p_{T} range is finite (both at low-p_{T} and at high-p_{T}) the fit Tsallis functions (Eq. 1.31) are used to compute the integrals.

The measured $d^2N/dydp_{\mathrm{T}}$ distributions in various y bins are shown in Fig. 7.2 for K^0_s and in Fig. 7.3 for Λ. It can be seen that most of the yield is in the non-perturbative low-p_{T} region.

The obtained dN/dy distributions are shown in Fig. 7.4 for K^0_s (left panel) and Λ (right panel). The numerical results are the following: $dN/dy|_{y=0} = 0.187 \pm 0.001 \pm 0.01$ and $0.314 \pm 0.002 \pm 0.02$ for K^0_s and $0.11 \pm 0.01 \pm 0.01$ and $0.171 \pm 0.002 \pm 0.01$ for Λ.

7.3 Discussion

The results presented in the previous two sections were obtained with the high efficiency method. They were found in agreement with the high purity analysis within 1.5σ for K^0_s and within less than 1σ for Λ. For the final CMS publication the two methods were combined into a single one taking the superior features of each analysis: 2D correction method and extra tracks due to the minimum bias

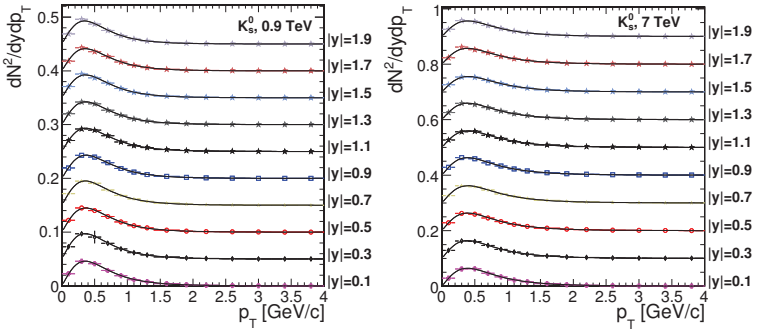


Figure 7.2: Measured $dN^2/dydp_\mathrm{T}$ distributions of K^0_s at 900 GeV (*left*) and at 7 TeV (*right*). Results are corrected to NSD collisions.

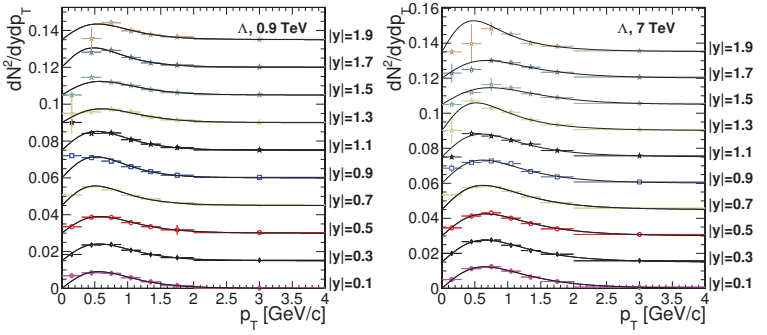


Figure 7.3: Measured $dN^2/dydp_\mathrm{T}$ distributions of Λ at 900 GeV (*left*) and at 7 TeV (*right*). Results are corrected to NSD collisions.

tracking from the high efficiency analysis, V0 finder and standard tracks from the high purity analysis [83]. While in the previous sections I only presented results from my method, here I use the final results of the combined method as the basis of discussion on strange hadron production at LHC.

The measured $\langle p_\mathrm{T} \rangle$ of K^0_s and Λ as a function of the centre-of-mass energy compared to the results of other experiments are shown in Fig. 7.5. The CMS

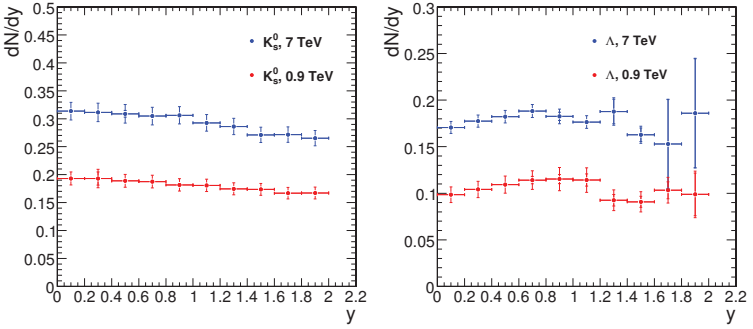


Figure 7.4: Measured dN/dy distributions of K^0_s (*left*) and Λ (*right*) at 0.9 and 7 TeV. The yields are obtained from the Tsallis fit functions.

measurements are for $|y| < 2$, the other experimental results are for various y intervals (since the $\langle p_T \rangle$ is weakly y dependent the figure is not to be taken as a precise comparison between the various results). The measured $\langle p_T \rangle$ increases with increasing particle mass and increasing centre-of-mass energy. The CMS results are in good agreement with those of ALICE [86], and provide strict constraints for phenomenological models. Note that ALICE reported the $\langle p_T \rangle$ for inelastic whereas the CMS result is for NSD collisions, but the $\langle p_T \rangle$ shows very little (if any) dependence on the collision type.

The p_T distributions in PYTHIA have different shapes compared to those in data, as shown in 7.6. It can be seen that the p_T distribution in data is softer than in PYTHIA. Although the various tunes show significant variations in the predicted distributions, neither of them are able to describe the observed shape. However, the measured and the predicted shapes are different, the $\langle p_T \rangle$ values are not so far away from each other at 7 TeV, while the PYTHIA values are remarkably lower at 0.9 TeV (the $\langle p_T \rangle$ values in various event generators are given in Table 7.1).

A comparison between the measured and predicted dN/dy as a function of y distributions for K^0_s (left panel) and Λ (right panel) are shown in Figure 7.7. This figure shows several interesting features. First, we notice that the observed densities of strange hadrons are significantly larger than the ones predicted by the various

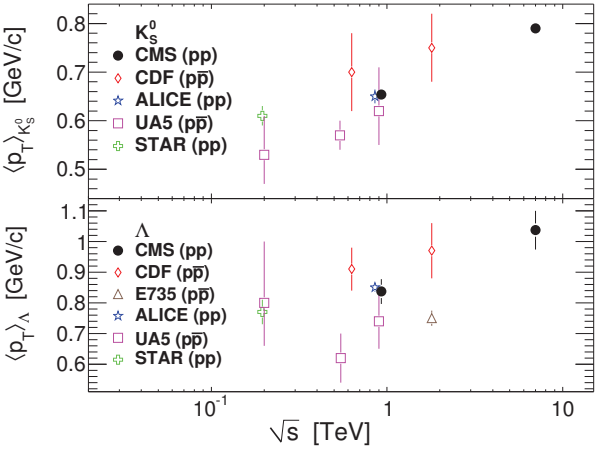


Figure 7.5: Average-p_T for charged hadrons, K^0_s (*top*) and Λ (*bottom*) as a function of the centre-of-mass energy. The CMS measurements are for $|y| < 2$. Also shown are results from UA5 [87, 88, 89, 90] ($p\bar{p}$ collisions covering $|y| < 2.5$ and $|y| < 2$ for K^0_s and Λ, respectively), E735 [91] ($p\bar{p}$ collisions using tracks with $-0.36 < \eta < 1.0$), CDF [92] ($p\bar{p}$ collisions covering $|\eta| < 1.0$), STAR [93] (pp collisions covering $|y| < 0.5$), and ALICE [86] (pp collisions covering $|y| < 0.75$ for K^0_s and Λ). Some points have been slightly offset from the true energy to improve visibility. The vertical bars indicate the statistical and systematic uncertainties (when available) summed in quadrature.

Particle	$\sqrt{s} = 0.9$ TeV $\langle p_T \rangle$ [GeV/c]	$\sqrt{s} = 7$ TeV $\langle p_T \rangle$ [GeV/c]
K^0_s, Data	$0.654 \pm 0.001 \pm 0.008$	$0.790 \pm 0.001 \pm 0.009$
K^0_s, PYTHIA6 D6T	0.581	0.753
K^0_s, PYTHIA8	0.550	0.713
K^0_s, PYTHIA6 P0	0.585	0.730
Λ, Data	$0.837 \pm 0.006 \pm 0.040$	$1.037 \pm 0.005 \pm 0.063$
Λ, PYTHIA6 D6T	0.756	1.064
Λ, PYTHIA8	0.666	0.933
Λ, PYTHIA6 P0	0.695	0.921

Table 7.1: Average-p_T results in data and in various PYTHIA simulations in 0.9 and 7 TeV non-single-diffractive proton-proton collisions.

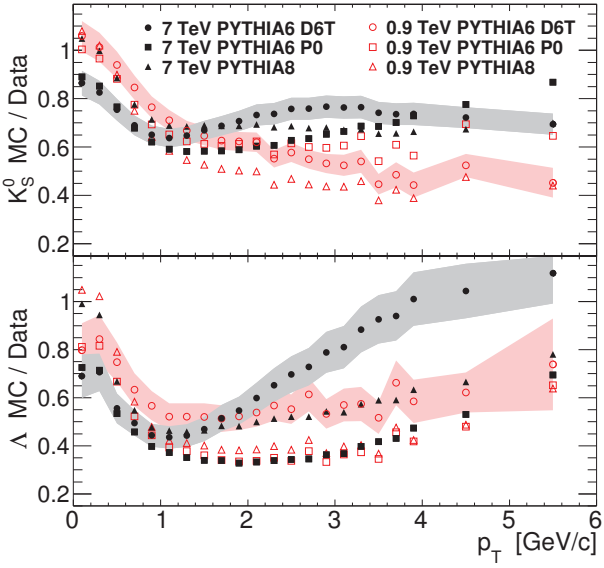


Figure 7.6: Ratio of the predicted p_T spectra of various PYTHIA tunes and the measured data for K^0_s and Λ at 0.9 (open symbols) and 7 TeV (filled symbols) NSD pp collisions. The uncertainties (in order to make the plot more transparent) are only shown for one of the predictions, D6T, at each energy.

PYTHIA tunes. This large deviation even takes the extreme form of the measured 0.9 TeV Λ production being well reproduced by the 7 TeV predictions! Second, it can be also seen that the measured yields as a function of y decrease for large rapidities. This decrease is not present in all tunes, it is only modeled by PYTHIA8 [19] and PYTHIA6 P0 [24] generators and tunes (out of the tested ones). Third, we see that there is a large increase in the measured yields of strange particles as the centre-of-mass energy increases from 0.9 to 7 TeV, which is not modeled by any of the PYTHIA tunes. The other related feature is that the deviation of the predicted and the measured production of strange particles is larger for the heavier particle, Λ. The numerical results on the deviation of the prediction of the standard CMS tune (D6T) and the measured yields, and the increase of particle yields from 0.9

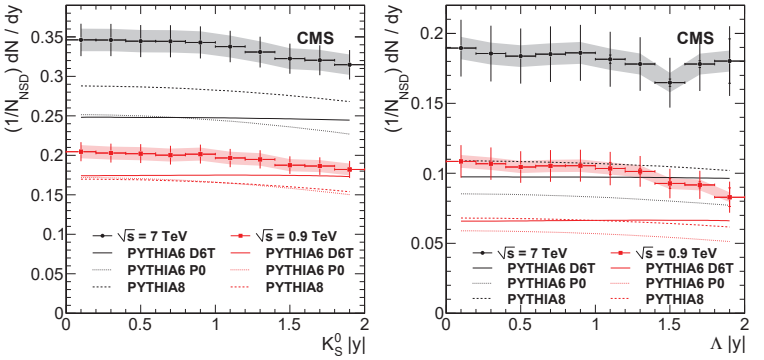


Figure 7.7: Measured K^0_s (*left*) and Λ (*right*) yields at 0.9 and 7 TeV centre-of-mass energy in comparison to three PYTHIA tunes [19, 24, 51].

	$\frac{dN}{dy}\|_{y=0}$(D6T)/$\frac{dN}{dy}\|_{y=0}$(Data)		$\frac{dN}{dy}\|_{y=0}$(7 TeV)/$\frac{dN}{dy}\|_{y=0}$(0.9 TeV)	
Particle	0.9 TeV	7 TeV	Data	D6T
K^0_s	0.86 ± 0.01 ± 0.07	0.72 ± 0.01 ± 0.05	1.69 ± 0.01 ± 0.06	1.42
Λ	0.61 ± 0.01 ± 0.07	0.51 ± 0.01 ± 0.06	1.75 ± 0.02 ± 0.08	1.48

Table 7.2: Comparisons of strange hadron densities between PYTHIA D6T and data at 0.9 and 7 TeV. In each column the first uncertainty represents the statistical whereas the second the systematic uncertainties.

to 7 TeV are presented in Table 7.2. The observed difference in K^0_s production at 0.9 TeV and 7 TeV are 14% and 28%, respectively, which are consistent with those observed for unidentified charged hadrons (see in Section 5), but not with deviation for Λ, which is significantly larger. The observed increase in K^0_s production from 0.9 to 7 TeV is 69% compared to the 42% predicted by D6T. The same underprediction is observed for Λ: we measure and increase of 75%, while the prediction of D6T is 48%.

The collision energy dependence of $dN/dy|_{y=0}$ is shown in Fig. 7.8. The earlier experimental data are taken from [87, 89, 94, 93, 86]. Note that the colliding systems (pp and $p\bar{p}$) as well as the measured collision types (inelastic and NSD) vary between the earlier experiments. Since the effect of different colliding systems on $dN/dy|_{y=0}$

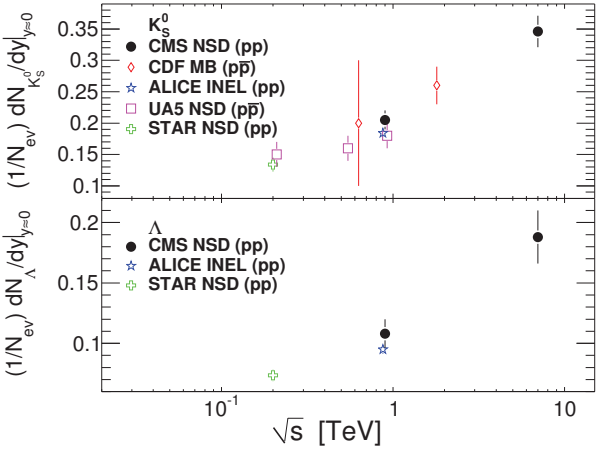


Figure 7.8: The central rapidity production rate for K^0_s (top) and Λ (bottom) as a function of the centre-of-mass energy. Results are also shown for UA5 [87, 89] ($p\bar{p}$), CDF [94] ($p\bar{p}$), STAR [93] (pp), and ALICE [86] (pp). The CMS, UA5, and STAR results are normalized to NSD events. The CDF results are normalized to events passing their trigger and event selection defined chiefly by activity in both sides of the detector, at least four tracks, and a primary vertex. The ALICE results are normalized to all inelastic events. Some points have been slightly offset from the true energy to improve visibility. The vertical bars indicate the statistical uncertainties for the UA5 and CDF results and the combined statistical and systematic uncertainties for the CMS, ALICE, and STAR results.

is only a few percent, whereas that of the collision type is ~ 15 percents, all of the various experimental results are in good agreement.

An interesting question is whether the observed ratio of strange hadrons changes as a function of collision energy. This could be the sign of novel production effects. The production ratios for K^0_s and Λ as a function of rapidity and transverse momentum are shown in Fig. 7.9. The rapidity distribution of the ratios are flat and do not show any dependence on collision energy (the fact that all 7 TeV points are consistently above the 0.9 TeV points might hint a very small dependence). The ratios as present in various PYTHIA tunes are also given. The fact that the ratios in simulation are smaller than the measured ones shows that the discrepancy between

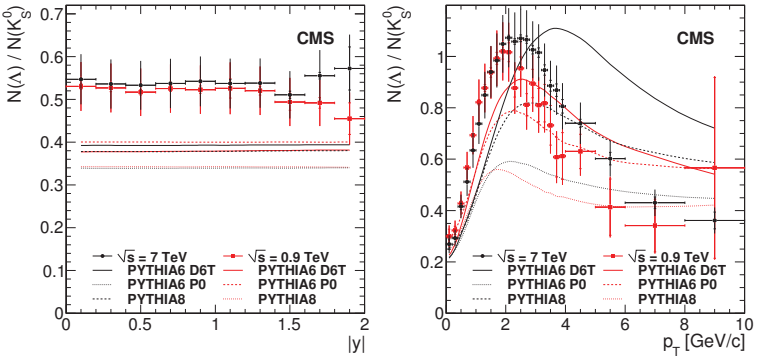


Figure 7.9: $N(\Lambda)/N(K^0_s)$ as a function of rapidity (*left*) and transverse momentum (*right*) in NSD events. The inner vertical error bars (when visible) show the statistical uncertainties, the outer ones show the statistical and all systematic uncertainties summed in quadrature. Results are also shown for three PYTHIA predictions at each centre-of-mass energy.

the predicted and measured K^0_s and Λ yields is larger for Λ. At the same time, the flatness of the distributions are well modelled by PYTHIA.

The ratios of the p_{T} distributions (right panel of Fig. 7.9) are also independent of the collision energy within uncertainties over the whole measured p_{T} range. Although, the data might indicate a small p_{T} dependence at high-p_{T}. The PYTHIA predictions are very close to each other at low-p_{T}, while they predict different high-p_{T} behaviours. Neither of the tunes can describe the shape of the measured ratio of the p_{T} distributions of Λ and K^0_s.

Summary

In this thesis I present three new analysis methods developed to measure the angular ($dN/d\eta$) and transverse momentum (dN/dp_T) distributions of unidentified charged hadrons ("pixel counting" and "minimum bias tracking") and identified strange particles ("high efficiency" method). Apart from having contributed to the development of the methods, I also performed two of them (pixel counting and high efficiency method) on the 0.9, 2.36, and 7 TeV centre-of-mass energy proton-proton collision data. The results of these measurements are also presented in the thesis.

The thesis starts with a brief theoretical overview on high energy proton-proton collisions. Then the CERN LHC and the CMS Experiment are introduced. This is followed by the description of the pixel counting and the minimum bias tracking methods used to measure the spectra of unidentified charged hadrons. The results of these methods on collision data are shown right after their description. Then the third new analysis method, the high efficiency method developed to measure the spectra of strange particles (K^0_s and $\Lambda/\bar{\Lambda}$), is described. Afterwards my results on collision data are presented.

My main results presented in this thesis are as follows:

- Three new analysis methods to measure the angular and transverse momentum distributions of unidentified charged hadrons and identified strange particles.

- Inclusive pseudorapidity distributions of unidentified charged hadrons at energies never measured before in proton-proton collisions. The distributions have only a slight pseudorapidity dependence in the measured region. The distribution at 0.9 TeV was found to be in a good agreement with earlier measurements in proton-antiproton collisions at the same energy. The measured charged hadron particle yields at 2.36 and 7 TeV were significantly larger than previously expected. Furthermore, the rate of the increase with collision energy also exceeded the expectations. None of the most popular event generators (PYTHIA, PHOJET) could reproduce the observed behaviour.

- Inclusive rapidity distributions of K^0_s and $\Lambda/\bar{\Lambda}$ at 0.9 and 7 TeV. The distributions are flat at midrapidity, they start to decrease at $y \approx 1.2$, which feature is only described by a few event generator models. The measured particle yields largely exceed the predictions of all tested event generators. The predictions of generators become worse with increasing energy collision energy and particle mass (for K^0_s being compatible with the deficiency seen for charged hadrons). The production ratio of K^0_s and $\Lambda/\bar{\Lambda}$ does not show any energy dependence.

- Inclusive transverse momentum distributions of K^0_s and $\Lambda/\bar{\Lambda}$ at 0.9 and 7 TeV. The distributions are accurately fit with the Tsallis function. The shape of the transverse momentum distributions were not well described by any of the tested event generators.

Appendix A
Fraction of diffractive dissociation in data

The pseudorapidity density distributions, $dN/d\eta$, of diffractive and non-diffractive events are very different as it can be seen in Fig. A.1. The uncertainty of the diffractive proportion in data propagates into the simulation based event selection correction. Thus, it is important to verify that the diffractive event fraction is not too different in data and simulation. This study was performed as follows.

In the first step an observable diffraction-sensitive observable is chosen. Then, the diffractive and non-diffractive distributions of this observable are taken from the simulation. The distribution measured in data is fit with a combination of the two shapes of the simulated distributions. The approximate fraction of diffraction in data is determined as the fraction of diffraction corresponding to the best fit.

Three observables were chosen to estimate the diffraction:

$$\sum_{HF^+}(E + p_z), \tag{A.1}$$

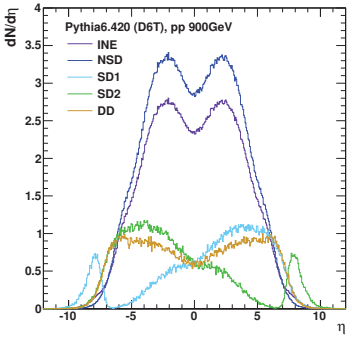


Figure A.1: Contributions of different event classes (inelastic, non-single-diffractive, single diffractive, and double diffractive) to the charged particle multiplicity in the PYTHIA D6T tune in pp collisions at 0.9 TeV.

$$\sum_{HF^-}(E - p_z), \tag{A.2}$$

$$\eta_{HF} = \frac{1}{2}\ln\left(\frac{\sum_{HF}(E + p_z)}{\sum_{HF}(E - p_z)}\right), \tag{A.3}$$

where E is the energy, $p_z = E\cos\theta$ is the longitudinal momentum component in the HF, and the sum runs over HF towers with energy greater than 3 GeV. This energy cut was chosen to reject the noise of HF (it is the same cut as the one applied during the selection of events for analysis, see Section 4.1). The study was performed with the PYTHIA and with the PHOJET event generators, respectively. They represent the extremes of simulations regarding the modelling of diffractive collisions. During this study the same event selection conditions were imposed as in the analyses described in the main body of this thesis except the HF coincidence, which would reject most of the diffractive events.

The above-mentioned diffraction-sensitive quantities are shown in Fig. A.2 and Fig. A.3 for the 0.9 TeV data. In Fig. A.2 the sharp peak in the $\sum_{HF^+}(E + p_z)$ distribution at 0 corresponds to events with zero HF towers (above 3 GeV). This dominantly happens in diffractive events. The long tail of the distribution reflects the multiplicity of the event. The best fits obtained with PYTHIA (left) and PHOJET (right) together with their two components (non-diffractive and diffractive) are also shown in the figures. The η_{HF} variable shown in Fig. A.3 also shows a peak around 0. This is composed by those events which have hits in both sides of the HF (+ and -). The two secondary peaks at $\sim \pm 4$ are due to diffractive events, as shown by the two components of the best fit. They only have hits in one side of the HF (the HF covers the $\sim 3 < |\eta| <\sim 5$ intervals).

The final estimation on the fraction of diffractive collisions is obtained by performing the previously-shown fits using various x-axis intervals. Since the whole η_{HF} interval is need for a meaningful η_{HF} fit, that is not varied. However, the $\sum_{HF^\pm}(E \pm p_z)$ can be fit using various x ranges: from 0 to 50, from 0 to 100, etc. This gives a whole set of results which are then averaged. The error of the average is computed as the RMS of the various fit results. The finally obtained results are

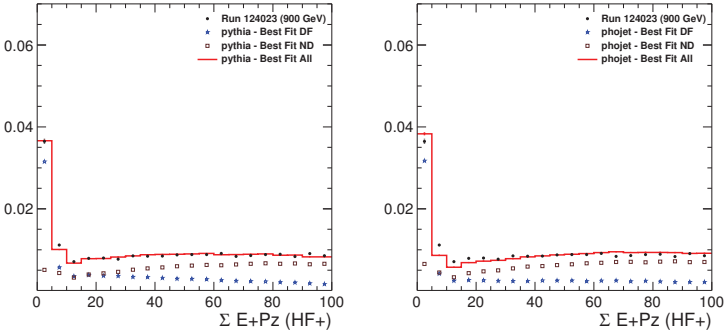


Figure A.2: Example fits of data into diffractive (DF) and non-diffractive (ND) components in terms of the $\sum_{HF^+}(E + p_z)$ quantity using the shapes of PYTHIA (*left*) and PHOJET (*right*) generators.

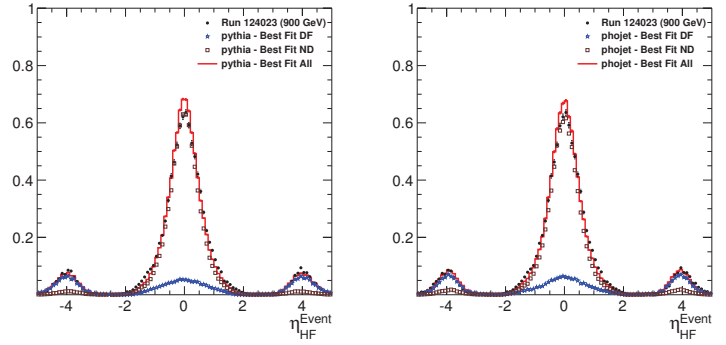


Figure A.3: Example fits of data into diffractive (DF) and non-diffractive (ND) components in terms of the η_{HF} quantity using the shapes of PYTHIA (*left*) and PHOJET (*right*) generators.

Sample		MC default	Data fit
0.9 TeV	PYTHIA	22.7%	$23.1 \pm 1.0\%$
	PHOJET	20.1%	$24.3 \pm 1.8\%$
2.36 TeV	PYTHIA	22.4%	$21.2 \pm 2.1\%$
	PHOJET	16.9%	$22.2 \pm 2.7\%$
7 TeV	PYTHIA	21.9%	$17.0 \pm 4.8\%$
	PHOJET	14.6%	$23.1 \pm 5.6\%$
	PYTHIA+PHOJET		$20.6 \pm 3.6\%$

Table A.1: Diffractive event fraction fit results for 0.9, 2.36, and 7 TeV. The errors are computed using various estimation methods and taking the RMS of the set of fit results.

summarized in Table A.1. At 7 TeV a third event shape was constructed combining the shapes from PYTHIA and PHOJET to better describe the data. The results with this hybrid function are also given is Table A.1.

This study supports the use of PYTHIA to calculate the MC corrections for the measurements I presented in this thesis. The diffractive event fraction of PYTHIA after event selection (except the HF coincidence) is in reasonable agreement with the fraction extracted from the data. Thus, for the measurement I used the PYTHIA value of diffractive events remaining in the data after the full event selection (including the HF coincidence).

Taking into account that the contribution of diffractive events (SD+DD) to the measured distributions are small, we may estimate a systematic uncertainty on the final results due to diffraction in the following way. First, the relative uncertainty in the measured fraction of diffractive events is $3.6/20.6 \sim 0.2$ (quoting the 7 TeV PHOJET+PYTHIA combined results). Since a larger SD or DD event fraction in data than simulation moves the final results (presented in the main body of this thesis) downwards, we may treat the uncertainties of SD and DD fractions together. The second point is that according to PYTHIA, after full event selection (including the HF coincidence) at 7 TeV 12.7% of the events are diffractive. Then the total uncertainty on the measured distributions due to the uncertainties in the fraction of diffractive events is $12.7\% \times 0.2 = 2.5\%$.

Appendix B
Properties of pixel clusters in data and simulation

B.1 Detector performance plots

Pixel clusters are used by every analysis presented in this thesis. It is important to validate the description of the simulation of their various features. The point is to make sure that the signals of the detector are understood. Small differences between the measured and predicted distributions might arise due to the difference in the properties of produced particles in the collision. I will examine this possibility wherever such deviation occurs.

The most important comparison plots using the innermost pixel layer were presented in the main text of the thesis, in Section 4.3.5. In this appendix a larger set of such performance plots are collected for the outermost layer.

Cluster length in the local y direction The local y direction (equivalent to the global z) is parallel to the beam line. It is important to check the description of the simulation of the cluster length, because both the pixel counting and the tracking analyses performs various selection cuts in this quantity (cluster selection, cluster shape filter). The distributions of cluster length can be seen in Fig. B.1 and Fig. B.2 for the third pixel barrel layer, at 0.9 and 7 TeV, respectively. The observed excess at low cluster sizes was already discussed in Section 4.3.5.

Cluster length in the local x direction In the tracking analysis clusters might be rejected based on the length in the local x direction by the cluster shape filter. Thus, it is important to check whether the data is understood based on the simulated distribution. Note that the magnetic field bends the trajectories in the local x direction. Hence, the p_{T} spectra are folded into these distributions. The distributions of the cluster length along the local x direction are shown in Fig. B.3 and Fig. B.4 for the third pixel barrel layer, at 0.9 and 7 TeV, respectively.

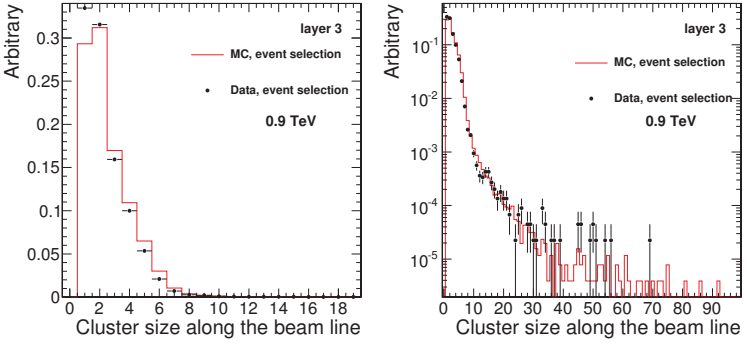


Figure B.1: Cluster size distributions along the local y direction in layer 3 after event selection at 0.9 TeV. Data are shown with black dots, the solid line shows the Monte Carlo expectation.

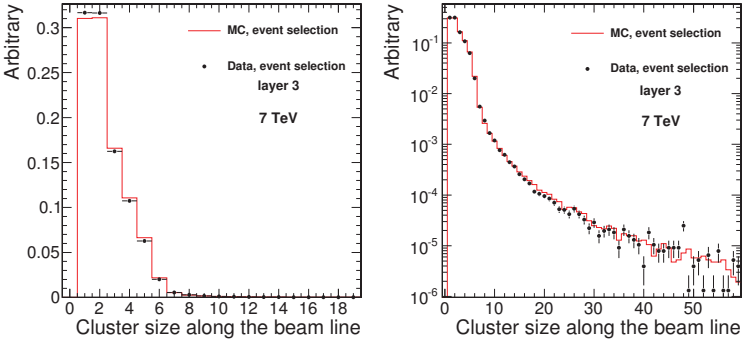


Figure B.2: Cluster size distributions along the local y direction in layer 3 after event selection at 7 TeV. Data are shown with black dots, the solid line shows the Monte Carlo expectation.

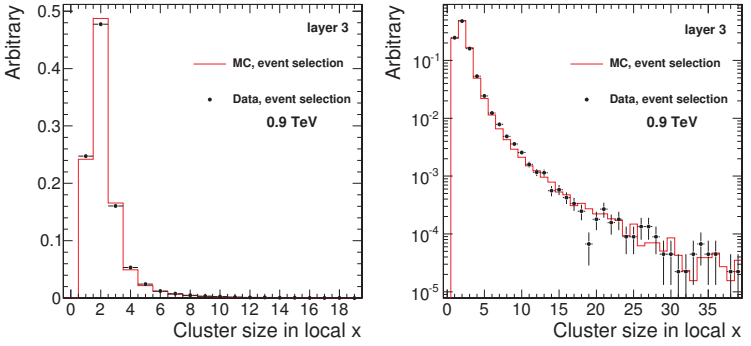


Figure B.3: Cluster size distributions along the local x direction in layer 3 after event selection at 0.9 TeV. Data are shown with black dots, the solid line shows the Monte Carlo expectation.

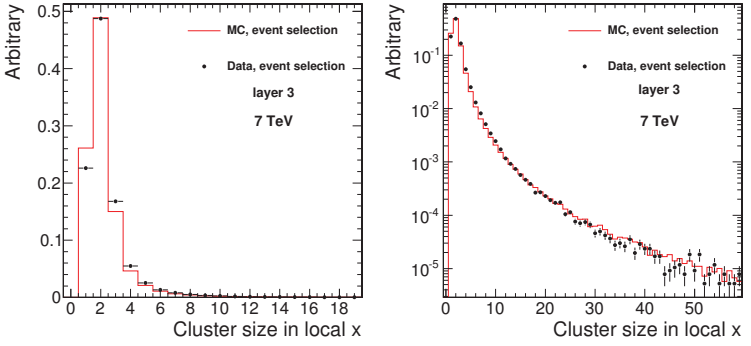


Figure B.4: Cluster size distributions along the local x direction in layer 3 after event selection at 7 TeV. Data are shown with black dots, the solid line shows the Monte Carlo expectation.

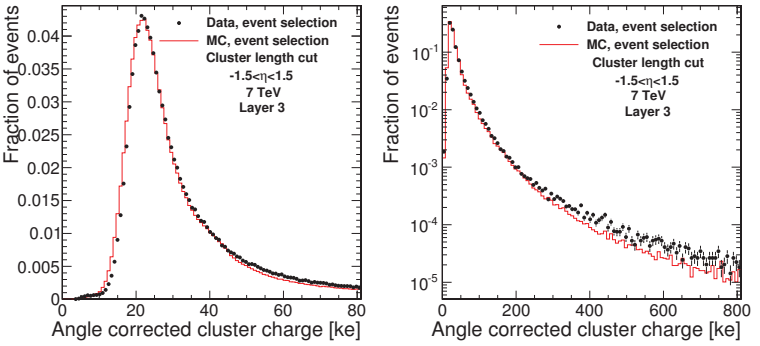


Figure B.5: Angle corrected cluster charge distributions in data and in simulation in layer 3 at 7 TeV.

Distribution of pixel cluster charge The angle corrected barrel pixel charge distributions using those clusters which passed the cluster size cut can be seen in Fig. B.5 for the third layer, at 7 TeV. The features of these distributions were already discussed in Section 4.3.5.

B.2 Event-by-event multiplicity distributions

The multiplicity distributions are expressed here as the distribution of the number of pixel clusters per events. This event-by-event multiplicity in the simulation is determined by the event generator models and parametrisations. Thus, a mismatch between simulation and the data in this quantity only carries information about the physics of the collision, but not about the accuracy of the detector simulation. A more detailed discussion on the importance of this distribution was given in Section 4.3.5.

The multiplicity distributions measured in data are compared the PYTHIA D6T tune at 0.9 and to the PYTHIA ATLAS tune at 7 TeV. These tunes produce widely different event-by-event multiplicities, but neither of them can describe the data perfectly. However, the ATLAS tune provides the high multiplicity events necessary for the computation of the multiplicity dependent correction factors. The

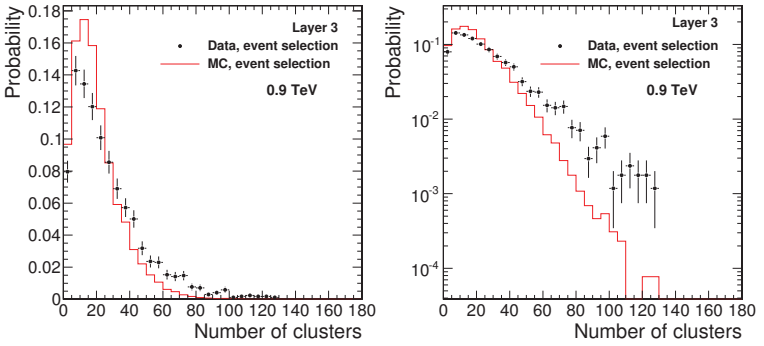


Figure B.6: Distribution of the number of clusters in layer 3 after event selection at 0.9 TeV. Data are shown with black dots, the solid line shows the Monte Carlo expectation.

multiplicity distributions are shown in Fig. B.6 and Fig. B.7 for the third pixel barrel layer at 0.9 and 7 TeV, respectively.

B.3 Sensitive detector surface

The "acceptance" corrections of the various analyses are directly taken from simulation. Thus it is important know what the fraction of unfunctional detector units in data is and whether such units are properly included in the simulation or not. By plotting the $\eta - \phi$ 2-dimensional distribution of the endpoints of tracklets one can map the pixel detector unit-by-unit. In Fig. B.8 the endpoint distribution of tracklets are shown in the first pixel barrel layer. It can be seen that there is a small area around $\eta \approx 0$ and $\phi \approx 1.6$ in the data which is not simulated. This extra insensitive area was taken into account at the time of data analysis by manually masking it in the simulated events. Although I only show the endpoint distribution for the first layer in the 2.36 TeV collisions, but it looks very similar for all the other layers at all collision energies.

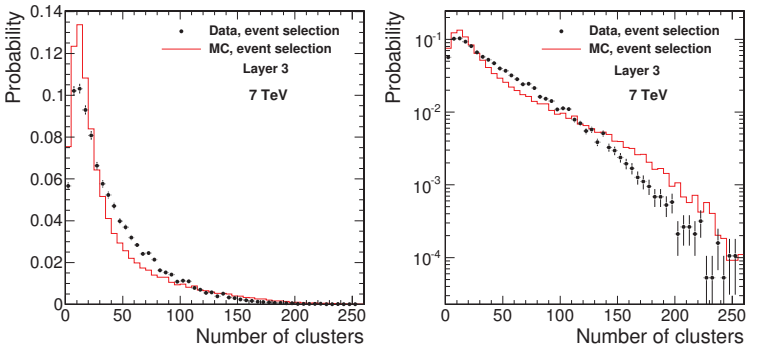


Figure B.7: Distribution of the number of clusters in layer 3 after event selection at 7 TeV. Data are shown with black dots, the solid line shows the Monte Carlo expectation.

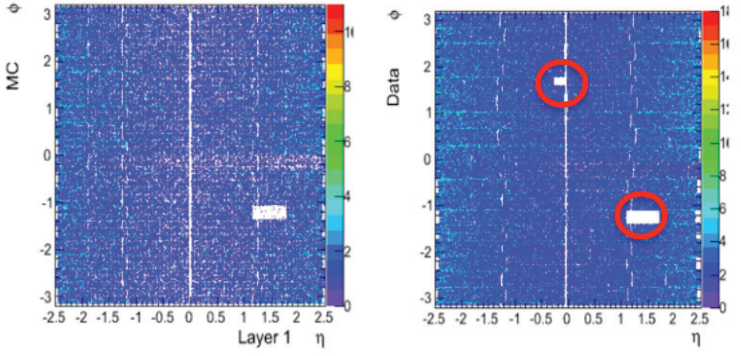


Figure B.8: Distribution of tracklet endpoints in the first pixel barrel layer at 2.36 TeV in the simulation (*left*) and in data (*right*). The insensitive detector areas appear as white rectangles. They are also noted with red circles on the left hand figure.

Acknowledgements

I would like to thank my parents for providing the support and background of the family that allowed me to continue working in the field of particle physics.

I would like to thank the support and the time of my supervisor, Gábor Veres, who helped me greatly in my research always sparing some time for discussion. I'm also grateful for him for the opportunity to spend two times half a year at CERN. One of these periods was right at the start of the LHC and the first data analysis marathon. It was a very intense period with a lot of things to remember.

I would like to thank Ferenc Siklér for the common work and vivid discussions, and also for the occasional funny burst outs and imitations which so tightly belong to his working style.

References

Note that in the list of references I also indicated internal CMS documentations, denoted as CMS Notes. Most of the CMS Notes have a much shortened public version, called CMS PAS, which are given along the CMS Notes in the references. The internal CMS Notes only have the real contributors as authors as opposed to the public CMS documents, which have the full collaboration as the author. Thus, it is difficult for a non-CMS member to disentangle the contribution of an individual to the success of the whole collaboration. However, there might be one quantity to measure this contribution, and that is the number of conference talks and posters an individual is awarded to represent the CMS Collaboration. Thus, I included in the list of references the talks and posters I gave on the topics of this thesis on international conferences. They are properly referenced in the main text of the thesis. According to the CMS Conference Committee, there were $\sim$ 100 talks given by CMS on international conferences in 2010. Taken into account that CMS has $\sim$ 3000 members, the expected number of talks is 0.03 per person assuming equal chances.

[1] CMS Collaboration. Observation of diffraction in proton-proton collisions at 900 and 2360 GeV centre-of-mass energies at the LHC. *CMS PAS FWD-10-001*, 2010.

[2] `http://nobelprize.org/nobel_prizes/physics/laureates/2004/`.

[3] D. J. Gross and Frank Wilczek. Asymptotically Free Gauge Theories. 1. *Phys. Rev.*, D8:3633–3652, 1973.

[4] CMS Collaboration. Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ TeV. *CMS PAS QCD-10-025*, 2011.

[5] CMS Collaboration. Measurement of the Inclusive Jet Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV. *CMS PAS QCD-10-011*, 2011.

[6] C. Alt et al. Inclusive production of charged poions in p+p collisions at 158 GeV/c beam momentum. *Eur. Phys. J. C*, 45:343, 2006.

[7] J. Adams et al. Identified particle distributions in p+p and Au+Au collisions at $\sqrt{s} = 200$ GeV. *Phys. Rev. Lett.*, 92:112301, 2004.

[8] J. Adams et al. Pion, kaon, proton and anti-proton transverse momentum distributions from p+p and d+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett. B*, 616:8, 2005.

[9] J. Adams et al. Identified hadron spectra at large transverse momentum in p+p and d+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett. B*, 637:161, 2006.

[10] G. Arnison et al. Transverse momentum spectra for charged particles at the CERN Proton Anti-Proton Collider. *Phys. Lett.*, B118:167, 1982.

[11] F. Abe et al. Transverse momentum distributions of charged particles produced in $\bar{p}p$ interactions at $\sqrt{s} = 630$ GeV and 1800 GeV. *Phys. Rev. Lett.*, 61:1819, 1988.

[12] C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *J. Statist. Phys.*, 52:479, 1988.

[13] C. Tsallis and E. P. Borges. Nonextensive statistical mechanics: Applications to nuclear and high energy physics. arXiv:cond-mat/0301521.

[14] G. Wilk and Z. Wlodarczyk. Power laws in elementary and heavy-ion collisions: A story of fluctuations and nonextensivity? *Eur. Phys. J*, A40:299, 2009.

[15] A. Jakovac T. S. Biro, G. Gyorgyi and G. Purcsel. A non-conventional description of quark matter. *J. Phys. G*, 31:S759–S763, 2005.

[16] S. M. Troshin and N. E. Tyurin. Energy dependence of average transverse momentum in hadron production due to collective effects. 2010. arXiv:1002.1554.

[17] E. Gotsman, E. Levin, and U. Maor. QCD motivated approach to soft interactions at high energy: inclusive production. *Phys. Rev.*, D81:051501, 2010.

[18] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[19] T. Sjöstrand, S. Mrenna and P. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852, 2008.

[20] Fritz W. Bopp, R. Engel, and J. Ranft. Rapidity gaps and the PHOJET Monte Carlo. 1998. arXiv:hep-ph/9803437.

[21] R. Engel, J. Ranft, and S. Roesler. Hard diffraction in hadron-hadron interactions and in photoproduction. *Phys. Rev.*, D52:1459, 1995.

[22] R. Bernhard et al. Proceedings of the First International Workshop on Multiple Partonic Interactions at the LHC (MPI08). 2010. arXiv:1003.4220.

[23] Andy Buckley, Hendrik Hoeth, Heiko Lacker, Holger Schulz, and Jan Eike von Seggern. Systematic event generator tuning for the LHC. *Eur. Phys. J.*, C65:331–357, 2010.

[24] Peter Z. Skands. The Perugia Tunes. 2009. arXiv:0905.3418.

[25] http://nobelprize.org/nobel_prizes/physics/laureates/1984/.

[26] https://edms.cern.ch/document/445830/5.

[27] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 03:S08004, 2008.

[28] ATLAS Collaboration. Measurement of the Inelastic Proton-Proton Cross-Section at $\sqrt{s} = 7$ TeV with the ATLAS Detector. arXiv:1104.0326.

[29] CMS Collaboration. The CMS tracker system project: technical design report. *CERN-LHCC-98-006*, 1998.

[30] CMS Collaboration. The CMS tracker: addendum to the technical design report. *CERN-LHCC-2000-016*, 2000.

[31] CMS Collaboration. Alignment of the CMS silicon tracker during commissioning with cosmic rays. *JINST*, 5:T03009, 2010.

[32] Ferenc Sikler and Krisztian Krajczar. CMS: minimum bias studies (conference talk), *MPI@LHC'08*, October 27-31, 2008, Perugia, Italy.

[33] Ferenc Sikler and Krisztian Krajczar. CMS: minimum bias studies. *DESY-PROC-2009-06*.

[34] Krisztian Krajczar. QCD Studies with CMS (conference poster), *PANIC 2008*, November 9-14, 2008, Eilat, Israel.

[35] Krisztian Krajczar. Charged particle spectra at CMS (conference talk), *EPS HEP 2009*, July 16-22, 2009, Krakow, Poland.

[36] Krisztian Krajczar. Measurement of charged particle spectra in pp collisions at CMS. *PoS(EPS-HEP 2009)062*.

[37] Krisztian Krajczar. p+p ütközések vizsgálata az LHC CMS detektorával (conference talk), *Magfizikus Találkozó 2009*, September 3-4, 2009, Jávorkút, Hungary.

[38] Krisztian Krajczar and Gabor Veres. Pseudorapidity distribution of charged hadrons in minimum bias p-p collisions at $\sqrt{s} = 14$ TeV. *CMS Note AN-2008/018 and CMS PAS QCD-08-004*, 2008.

[39] Yen-Jie Lee et al. Study of Charged Hadron Multiplicity in Minimum Bias pp Collisions at $\sqrt{s} = 900$ GeV and 10 TeV. *CMS Note AN-2009/076 and CMS PAS QCD-09-002*, 2009.

[40] Ferenc Sikler and Krisztian Krajczar. Transverse momentum and pseudorapidity distribution of charged hadrons in proton-proton collisions at $\sqrt{s} = 7$ TeV. *CMS PAS QCD-09-008*, 2009. This CMS PAS was never made public.

[41] Ferenc Sikler and Krisztian Krajczar. Measurement of charged hadron spectra in proton-proton collisions at $\sqrt{s} = 14$ TeV. *CMS Note AN-2007/021 and CMS PAS QCD-07-001*, 2007.

[42] T. Alexopoulos et al. Multiplicity dependence of the transverse momentum spectrum for centrally produced hadrons in antiproton-proton collisions at $\sqrt{s}$ = 1.8 TeV. *Phys. Rev. Lett.*, 60:1622, 1988.

[43] G. J. Alner et al. Scaling of pseudorapidity distributions at c.m. energies up to 0.9 TeV. *Z. Phys.*, C33:1, 1986.

[44] C. Albajar et al. A Study of the General Characteristics of Proton-Antiproton Collisions at 0.2 to 0.9 TeV. *Nucl. Phys.*, B335:261, 1990.

[45] F. Abe et al. Pseudorapidity distributions of charged particles produced in $p\bar{p}$ interactions at $\sqrt{s} = 630$ GeV and 1800 GeV. *Phys. Rev.*, D41:2330, 1990.

[46] K. Aamodt et al. Charged-particle multiplicity measurement in proton-proton collisions at $\sqrt{s} = 0.9$ and 2.36 TeV with ALICE at LHC. *Eur. Phys. J.*, C68:89–108, 2010.

[47] B. I. Abelev et al. Systematic measurements of identified particle spectra in p+p, d+Au and Au+Au collisions at the STAR detector. *Phys. Rev.*, C79:034909, 2009.

[48] Ferenc Sikler. Study of clustering methods to improve primary vertex finding for collider detectors. *Nucl. Instrum. Meth.*, A621:526, 2010.

[49] C. Buttar A. Moraes and I. Dawson. Prediction for minimum bias and the underlying event at LHC energies. *Eur. Phys. J.*, C50:435, 2007.

[50] CMS Collaboration. Studies of tracker material. *CMS-PAS-TRK-10-003*, 2010.

[51] Paolo Bartalini and Livio Fano (ed.). Multiple Parton Interactions at the LHC. *DESY-PROC*, 06, 2009.

[52] R. Fruhwirt. Application of kalman filtering on track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987.

[53] Krisztian Krajczar, Gabor Veres, Ferenc Sikler et al. Transverse momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 900$ GeV and 2.36 TeV. *CMS Note AN-2009/182 and CMS PAS QCD-09-010*, 2009.

[54] Krisztian Krajczar, Gabor Veres, Ferenc Sikler et al. Transverse momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV. *CMS Note AN-2010/069 and CMS PAS QCD-10-006*, 2010.

[55] CMS Collaboration. Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 0.9$ and 2.36 TeV. *JHEP*, 02:041, 2010.

[56] CMS Collaboration. Transverse-momentum and pseudorapiditydistributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV. *PRL*, 105:022002, 2010.

[57] Krisztian Krajczar. QCD Physics from CMS (conference talk), *DIS 2010*, April 19-23, 2010, Florence, Italy.

[58] Krisztian Krajczar. QCD Physics from CMS. *PoS(DIS 2010)117*.

[59] Krisztian Krajczar. Töltött hadronok spektrumainak mérése 0,9, 2,36 és 7 TeV ütközési energiákon a CMS detektorral (conference talk), *Fizikus Vándorgyűlés 2010*, August 24-27, 2010, Pécs, Hungary.

[60] UA5 Collaboration. Scaling of pseudorapidity distributions at c.m. energies up to 0.9 TeV. *Z. Phys*, C33, 1986.

[61] A. M. Rossi et al. Experimental Study of the Energy Dependence in Proton Proton Inclusive Reactions. *Nucl. Phys.*, B84:269, 1975.

[62] J. Whitmore. Experimental Results on Strong Interactions in the NAL Hydrogen Bubble Chamber. *Phys. Rept.*, 10:273, 1974.

[63] W. Thome et al. Charged Particle Multiplicity Distributions in p p Collisions at ISR Energies. *Nucl. Phys.*, B129:365, 1977.

[64] Rachid Nouicer et al. Pseudorapidity distributions of charged particles in d+Au and p+p collisions at $\sqrt{s_{\rm NN}} = 200$ GeV. *J. Phys.*, G30:S1133, 2004.

[65] CMS Collaboration. Charged particle transverse momentum spectra in pp collisions at sqrt(s) = 0.9 and 7 TeV. 2011. arXiv:1104.3547.

[66] F. Arleo, D. dEnterria, and A. S. Yoon. Single-inclusive production of large-pT charged particles in hadronic collisions at TeV energies and perturbative QCD predictions. *JHEP*, 06:035, 2010.

[67] ALTAS Collaboration. Charged-particle multiplicities in pp interactions at $\sqrt{s} = 900$ GeV measured with the ATLAS detector at the LHC. *Phys. Lett.*, B688:21, 2010.

[68] `http://lpcc.web.cern.ch/lpcc/`.

[69] CMS Collaboration. Pseudorapidity distributions of charged particles in pp collisions at $\sqrt{s} = 7$ TeV and 0.9 TeV with at least one central charged particle. *CMS PAS QCD-10-024*, 2010.

[70] `http://www.phys.ufl.edu/~rfield/cdf/LPCC2_RickField_6-17-11.pdf`.

[71] CMS Collaboration. Charged particle multiplicities in pp interactions at $\sqrt{s}$ = 0.9, 2.36, and 7 TeV. *JHEP*, 01:079, 2011.

[72] CMS Collaboration. Observation of long-range, near-side angular correlations in proton-proton collisions at the LHC. *JHEP*, 09:091, 2010.

[73] STAR Collaboration. Experimental and Theoretical Challenges in the Search for the Quark Gluon Plasma: The STAR Collaborations Critical Assessment of the Evidence from RHIC Collisions. *Nucl. Phys.*, A757:102, 2005.

[74] `http://qm2011.in2p3.fr/node/12`.

[75] `https://indico.cern.ch/contributionDisplay.py?contribId=438&sessionId=45&confId=30248`.

[76] W.-T. Deng, X.-N. Wang, and R. Xu. Gluon shadowing and hadron production in heavy-ion collisions at LHC. arXiv:1011.5907.

[77] J. L. Albacete and A. Dumitru. A model for gluon production in heavy-ion collisions at the LHC with rcBK unintegrated gluon densities. arXiv:1011.5161.

[78] F. Bopp et al. Inclusive distributions at the LHC as predicted from the DPMJET-III model with chain fusion. 2007. arXiv:0706.3875. Interpolated between 2.0 and 5.5 TeV values.

[79] Krisztian Krajczar. Spectra of identified neutral hadrons at 0.9 and 7 TeV using decay topology. *CMS Note AN-2010/144 and CMS PAS QCD-09-007*, 2010. Common CMS PAS with the *high purity* method.

[80] Kevin Stenson et al. Measurement of the production of K^0_S, Λ^0, and Ξ^- particles from proton-proton collisions at center-of-mass energies 0.9 and 7 TeV. *CMS Note AN-2010/084 and CMS PAS QCD-09-007*, 2010. Common CMS PAS with the *high efficiency* method.

[81] H. Yamamoto. dE/dx particle identification for collider detectors. arXiv:hep-ex/9912024.

[82] Wolfgang Adam. Track and vertex reconstruction in CMS. *Nucl. Instrum. Meth.*, A582:781–784, 2007.

[83] CMS Collaboration. Strange particle production in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV. *JHEP*, 05:064, 2011.

[84] Krisztian Krajczar. Spectra of weakly decaying indetified particles at 0.9 and 7 TeV with the CMS detector (conference poster), *HCBM 2010*, August 15-20, 2010, Budapest, Hungary.

[85] Krisztian Krajczar. Spectra of weakly decaying identified particles at 0.9 and 7 TeV with the CMS detector. *EPJ Web of Conferences*, 13:03003, 2011.

[86] ALICE Collaboration. Strange particle production in proton-proton collisions at $\sqrt{s} = 0.9$ TeV with ALICE at the LHC. 2010. arXiv:1012.3257.

[87] UA5 Collaboration. Kaon production in $p\bar{p}$ reactions at a centre-of-mass energy of 540 GeV. *Nucl. Phys.*, B258:505, 1985.

[88] G. J. Alner et al. UA5: A general study of proton-antiproton physics at $\sqrt{s} =$ 546 GeV. *Phys. Rept.*, 154:247, 1987.

[89] R. E. Ansorge et al. Kaon production in $\bar{p}p$ interactions at c.m. energies from 200 to 900 GeV. *Z. Phys.*, C41:179, 1988.

[90] R. E. Ansorge et al. Hyperon Production at 200 and 900 GeV C.M. Energy. *Nucl. Phys.*, B328:36, 1989.

[91] T. Alexopoulos et al. Hyperon production from $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev.*, D46:2773, 1992.

[92] CDF Collaboration. K^0_S and Λ^0 Production Studies at $\sqrt{s} = 1800$ and 630 GeV. *Phys. Rev.*, D72:052001, 2005.

[93] STAR Collaboration. Strange particle production in $p + p$ collisions at $\sqrt{s} = 200$ GeV. *Phys. Rev.*, C75:064901, 2007.

[94] F. Abe et al. K^0_S production in $\bar{p}p$ interactions at $\sqrt{s} = 630$ GeV and 1800 GeV. *Phys. Rev.*, D40:3791, 1989.

Summary

In this thesis I present three new analysis methods developed to measure the angular ($dN/d\eta$) and transverse momentum (dN/dp_T) distributions of unidentified charged hadrons ("pixel counting" and "minimum bias tracking") and identified strange particles ("high efficiency" method). Apart from having contributed to the development of the methods, I also performed two of them (pixel counting and high efficiency method) on the 0.9, 2.36, and 7 TeV centre-of-mass energy proton-proton collision data. The results of these measurements are also presented in the thesis.

The thesis starts with a brief theoretical overview on high energy proton-proton collisions. Then the CERN LHC and the CMS Experiment are introduced. This is followed by the description of the pixel counting and the minimum bias tracking methods used to measure the spectra of unidentified charged hadrons. The results of these methods on collision data are shown right after their description. Then the third new analysis method, the high efficiency method developed to measure the spectra of strange particles (K^0_s and $\Lambda/\bar{\Lambda}$), is described. Afterwards my results on collision data are presented.

My main results presented in this thesis are as follows:

- Three new analysis methods to measure the angular and transverse momentum distributions of unidentified charged hadrons and identified strange particles.
- Inclusive pseudorapidity distributions of unidentified charged hadrons at energies never measured before in proton-proton collisions. The distributions have only a slight pseudorapidity dependence in the measured region. The distribution at 0.9 TeV was found to be in a good agreement with earlier measurements in proton-antiproton collisions at the same energy. The measured charged hadron particle yields at 2.36 and 7 TeV were significantly larger than previously expected. Furthermore, the rate of the increase with collision energy also exceeded the expectations. None of the most popular event generators (PYTHIA, PHOJET) could reproduce the observed behaviour.
- Inclusive rapidity distributions of K^0_s and $\Lambda/\bar{\Lambda}$ at 0.9 and 7 TeV. The distributions are flat at midrapidity, they start to decrease at $y \approx 1.2$, which feature is only described by a few event generator models. The measured particle yields largely exceed the predictions of all tested event generators. The predictions of generators become worse with increasing energy collision energy and particle mass (for K^0_s being compatible with the deficiency seen for charged hadrons). The production ratio of K^0_s and $\Lambda/\bar{\Lambda}$ does not show any energy dependence.
- Inclusive transverse momentum distributions of K^0_s and $\Lambda/\bar{\Lambda}$ at 0.9 and 7 TeV. The distributions are accurately fit with the Tsallis function. The shape of the transverse momentum distributions were not well described by any of the tested event generators.

Összefoglaló

A doktori disszertációban bemutatok három új adatkiértékelési módszert, melyeket az azonosítatlan töltött hadronok („pixel számlálás" és „minimum bias tracking" módszer) és azonosított ritka részecskék („nagy hatásfokú" módszer) szög- ($dN/d\eta$) és transzverzális impulzus-eloszlásának (dN/dp_{T}) mérésére fejlesztettünk ki. A fejlesztéshez való hozzájáruláson kívül a három módszer közül kettőt (a pixel számlálás és nagy hatásfok módszerét) alkalmaztam is a 0,9, 2,36 és 7 TeV tömegközépponti ütközési energiájú proton-proton ütközésekben. A disszertációban ezen mérési eredményeket is bemutatom.

A disszertáció elején rövid áttekintést adok a nagyenergiájú proton-proton ütközések elméletéről. Ezt követi a CERN LHC és a CMS-kísérlet bemutatása. Ezután az azonosítatlan töltött hadronok spektrumainak méréséhez használt pixel számlálás és minimum bias tracking módszer kerül bemutatásra, majd ezen módszerek ütközési adatokon elért eredményeit ismertetem. A harmadik adatkiértékelési módszer, a ritka részecskék (itt K^0_s és $\Lambda/\bar{\Lambda}$) mérésére kifejlesztett nagy hatásfokú módszert ezt követően tárgyalom. A módszer leírását az ütközési adatokon elért eredményeinek bemutatása követi.

A disszertációban bemutott főbb eredményeim a következők:

- Három új adatkiértékelési módszer az azonosítatlan töltött hadronok és azonosított ritka részecskék szög- és transzverzális impulzus-eloszlásának mérésére.

- Azonosítatlan töltött hadronok inkluzív pszeudorapiditás-eloszlásai korábban sosem vizsgált energiákon proton-proton ütközésekben. Az eloszlások csupán gyenge pszeudorapiditás-függést mutatnak a mérési tartományban. A 0,9 TeV-en mért eloszlások jó egyezésben vannnak a korábbi ugyanilyen energiájú proton-antiproton ütközésekben mértekkel. A 2,36 és 7 TeV-en mért töltött-hadron produkció a várakozásokat jelentősen meghaladta. A legnépszerűbb eseménygenerátorok (PYTHIA, PHOJET) nem tudták reprodukálni a megfigyelt viselkedést.

- K^0_s és $\Lambda/\bar{\Lambda}$ inkluzív rapiditás-eloszlásai 0,9 és 7 TeV-en. Az eloszlások középrapiditásnál laposak, de $y \approx 1.2$ környékén csökkeni kezdenek, amelyet csak néhány eseménygenerátor tud leírni. A mért részecskehozamok jelentősen meghaladják az összes vizsgált eseménygenerátor jóslatait. A jóslatok növekvő ütközési energiára és tömegre rosszabbak (K^0_s esetés a különbség még egyezik a töltött hadronoknál mérttel). A K^0_s és $\Lambda/\bar{\Lambda}$ keletkezési aránya energiafüggetlen.

- K^0_s és $\Lambda/\bar{\Lambda}$ inkluzív transzverzális impulzus-eloszlása 0,9 és 7 TeV-en. Az eloszlások jól illeszthetők Tsallis-függvénnyel. A transzverzális impulzus-eloszlások alakját egyik vizsgált eseménygenerátor sem tudta jól leírni.